THE MENTAL TIMELINE IN DISCOURSE ORGANIZATION AND PROCESSING

by

CHOONKYU LEE

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Karin Stromswold

and approved by

____

____

____

____

New Brunswick, New Jersey

October, 2012

ABSTRACT OF THE DISSERTATION

The mental timeline in discourse organization and processing

By CHOONKYU LEE

Dissertation Director:

Karin Stromswold

Early language research has revealed important insights into the building blocks of language, such as morphosyntactic features and rules and truth-conditions of sentences. Once we situate language in real-life use, however, a wide range of factors come into play. Language interacts not only with the surrounding linguistic context but also with the situational context, our mental representation of content, and our background knowledge. The discourse-level interaction among linguistic and extralinguistic factors is relevant to both sides of the communication – the speaker, in choosing and organizing linguistic expressions, and the listener, in selecting among different possible structures and meanings for the linguistic input.

The question I address in this dissertation is 'how we keep track of time when we use language.' My specific interests are (1) whether story time in narrative discourse is one of the critical dimensions that are dynamically updated as discourse progresses, and (2) how fine-grained our time representation is for discourse – whether it is simply an ordering of temporal points and intervals for the events and states described in the discourse, or a timeline where duration is preserved in greater detail.

In order to elaborate on these issues, I discuss results from my narrative production experiment and my narrative comprehension experiment. In the production study, based on wordless picture books, two kinds of linguistic expressions were found much more frequently after longer intervals in story time compared to shorter intervals: (1) explicit temporal marking with lexical or phrasal markers of topic time (e.g., *when*, *the next morning*, etc.); and (2) proper names in referring back to previously mentioned characters. In the comprehension study, based on short “two-minute mysteries,” longer duration in temporal adverbials in the stories tended to lead to longer reading times.

I conclude that magnitudes such as duration in story content are preserved in our linguistic encoding and have observable impact on our linguistic decoding, and extend the situation-model framework of discourse comprehension (van Dijk & Kintsch, 1983; Zwaan, 1999) to discourse production. My findings thus support an account of communication as alignment of situation models (Pickering & Garrod, 2006).

Acknowledgement

In a way, this section is the most “fun” part of a dissertation to write, although retrieving from my memory all the people who deserve thanks was not much easier than writing the main chapters. (The task, I’m afraid, may be even more incomplete than my main work.) A great number of people have helped me with my dissertation directly and indirectly.

First, I thank my advisor and mentor, Karin Stromswold. My graduate work at Rutgers has been a true odyssey, with everything that happened for the two of us combined over the years. Among the memories too numerous to recount, I thank her particularly for two things: I could count on regaining a sense of direction and motivation after each meeting with her, and she half-seriously offered to adopt me if necessary to prevent me from seeing battle with North Korea. She directly contributed to all parts of the dissertation with helpful suggestions for revision.

I also thank my committee members, Anthony Gillies, Julien Musolino, and Matthew Stone. Colleagues from other institutions were impressed when I mentioned my committee, and it has been a true privilege to work with such experts in different but related fields.

The interdisciplinary strength of Rutgers also allowed me to collaborate with and learn much from Smaranda Muresan, who contributed directly to Chapter 4 on Natural Language Processing tools for analysis of referring expressions in discourse. I thank her and Nina Wacholder for providing opportunities to learn about NLP through colloquia and reading group discussions, and for providing encouragement in their Language and Information course.

My colleagues in the Language Acquisition and Language Processing Lab have also offered invaluable help. Gaurav Kharkwal wrote most of the codes for the experiments and contributed to the experimental design in the process as well. Nora Isacoff and Gaurav provided important feedback about the stimuli and statistical analysis. I have also been blessed to have wonderful undergraduate research assistants and volunteers, Emily Schwarz, Earnestine Qiu, and Marina Donnelly, who stuck it out through hours and hours of data coding and data collection. I also thank other current and former members of the lab, including Tina Hou and Diane Molnar.

The audience at the LSA 2012 Annual Meeting, especially the special session in memory of Ellen Prince, provided very helpful comments. My special thanks go to Barbara Abbott, Robert Botne, Shobhana Chelliah, Sharon Cote, Jeruen Dery, Sophia Malamud, Eleni Miltsakaki, Livia Polanyi, and the anonymous reviewers. The participants in the NAACL-HLT 2012 provided helpful suggestions as well. Special thanks go to Michael Flor, Anna Kazantseva, Inderjeet Mani, and the anonymous reviewers. The experience at these conferences was a great honor for me, and also made me believe in cognitive clones.

Back to Rutgers, I'd like to thank those who have been a constant influence, enriching my life with their support and intellect. In Psychology and Cognitive Science, I thank Rochel Gelman and the instructors of the courses I took for shaping my thought and providing the pointers. Rochel has been constantly encouraging and goading in my study of number words and magnitude representation. I also thank fellow graduate students and alumni, including Sydney Levine, Marta Suárez, and John Wilder, who discussed my work with me at various stages. In Linguistics, Maria Bittner helped me develop my understanding of dynamic semantics, especially with regard to time in discourse, through her stimulating course and seminar. Roger Schwarzschild was always eager to hear about my work and career (and helped me greatly with my preparation for a job interview). Jane Grimshaw was my go-to linguist for syntax and number words, and she also offered helpful career advice. I also thank Veneeta Dayal for her Semantics course, Mark Baker for being a big factor in my decision to come to Rutgers on the preview day for prospective students, and graduate students and alumni of Linguistics, including Daniel Altshuler, Vandana Bajaj, Carlos Fasola, Todor Koev, and Sarah Murray, as well as the SALT co-organizers – SALT 2011 at Rutgers was a lot of fun! In the School of Communication and Information, I would like to thank Soo-Yeon Hwang, Michael Cole, and Stephen DiDomenico.

I thank my former and current colleagues from Korea in Psychology, Sung-Ho Kim, Seha Kim, and Haewon Yoon; other fellow Korean graduate students, postdocs, and alumni of Rutgers, including Sangho Oh, Lee-Sun Choi, Mee Sook Kim, Seol Ki, Minkyoung Yoo, and Michelle; friends from the Graduate Student Association; former and current NYU Linguistics folks; friends from the Plangere and Livingston Writing Centers; Ford Hall and Buell Apartments neighbors; friends from New Brunswick and New York, including Melanie Kelliher and Yoonhee Kang; Korean scientists and engineers' local soccer team members and local church members; and friends from Korea – Youngjoo Lee, Shinae Kang, Won, Youngho, Hyewon, Jae-Hwa, Jin, and my brothers-in-arms, who carry truly indelible memories of joy, loss, fear, regret, patience – life! Special thanks go to those who helped me early in my graduate career at Rutgers: Euichul Jung, Jongjin Jung, Baik Hoh, Eun Kyung Lee, Hyunjoo Kim, and Seunghun Lee.

I thank Graduate Vice Chair George Wagner and Cognitive Area Coordinator Jacob Feldman for allowing my smooth transition from graduate school to the army and back to graduate school. Also, the administrative staff of Psychology and RuCCS are the friendliest one can imagine. They actually made bureaucracy bearable. I thank, in particular, Anne Sokolowski, Jo'Ann Meli, Sue Cosentino, Donna Tomaselli, Dianne McMillion, John Dixon, and Larry Pishioneri, as well as former staff members Patty McGuire and Dianna Richter. Despite how stressful their jobs must be, they never let it show. I also thank the staff of the Center for International Faculty and Student Services and the Graduate School – New Brunswick, particularly the GradFund staff.

In addition, I'm grateful to the Psychology Building staff, who are always the first to come in to keep the building clean for the department. They also taught me some Spanish: I will always remember *sábado y domingo* are something to look forward to. I'm also grateful to the baristas of the Starbucks on George Street and the Au Bon Pain on College Avenue for keeping me going. They've become pretty good at predicting what I want.

Last but not least, I thank my sister, Suh-Kyung Lee, and my family for their constant support through thick and thin.

Some of the chapters in this dissertation are based on earlier manuscripts and abstracts. Chapter 2 is based on Lee, Kharkwal, and Stromswold (2012); Chapter 3 is based on Lee (2012) and Lee and Stromswold (2012); Chapter 4 is based on Lee, Muresan, and Stromswold (2012). I thank the Linguistic Society of America and the North American Chapter of the Association for Computational Linguistics.

Sample stimuli in Figures 1 and 2 are from Mercer Mayer's picture books (see List of Illustrations for the copyright notice).

I dedicate this work to my parents – Chungmin Lee, a linguist at Seoul National University, and Hyekyung Park, a domestic engineer. One of my father's favorite movie characters is Mozart Sr. in the movie *Amadeus*, and I am sympathetic to Mozart Jr. for the pressure he often felt from his father – if only I had Amadeus' talent too! I can't imagine even approaching the level of constant, sustained passion my father has for his work. Every farmer who owns a donkey beats it, and every son who has a professionally accomplished father lives in his shadow. I resisted his penchant for linguistics, and became a psycho-linguist. (My sister resisted in her own way, and studied speech and hearing sciences. Rebelliousness actually runs in the family: My dad defied my grandfather, who told him to study law, by deciding to study the laws of language instead.) My mother is yet her own kind of language expert with a degree in language education. Of course, she is much more than that (n.b., my master's thesis).

Table of Contents

List of Tables

List of Illustrations

* From FROG, WHERE ARE YOU? by Mercer Mayer, copyright (c) 1969 by Mercer Mayer. Used by permission of Dial Books for Young Readers, a division of Penguin Group (USA) Inc.

** From FROG GOES TO DINNER by Mercer Mayer, copyright (c) 1974, renewed 2002 by Mercer Mayer. Used by permission of Dial Books for Young Readers, a division of Penguin Group (USA) Inc.

Chapter 1

Introduction

Producing a clear and coherent narrative discourse involves the use of cohesive devices to reflect continuity in the content (e.g., pronouns for maintenance of reference) and markers of discourse boundaries to indicate breaks or shifts in the content (e.g., temporal adverbial phrases indicating passage of time). Understanding a narrative requires keeping track of various dimensions of the content dynamically to form a coherent mental representation. We need to keep track of the topic time (Klein, 1994) of the discourse, the setting, and the main characters, among other things (Magliano, Zwaan, & Graesser, 1999).

The importance of time as a core dimension of discourse meaning has been noted across disciplines, from psycholinguistics (e.g., Zwaan, 1999) to semantics (e.g., Bittner, 2012). Languages such as English have obligatory tense marking and frequent aspectual markers in main clauses for a rich specification of temporal locations and relations, and these markers relate eventualities temporally within and across sentences (e.g., Partee, 1973; Webber, 1988; Moens & Steedman, 1988). So-called ‘tenseless’ languages (such as Yucatec Maya, Bohnemeyer, 2009; Kalaallisut, Bittner, 2011; Paraguayan Guaraní, Tonhauser, 2011) also express rich temporal relations through aspectual or modal markers. The tasks of discourse production and comprehension thus require the narrator to plan, and the addressee to be aware of, changes in topic time globally throughout a discourse.

1.1. The Mental Timeline

Time is unlike many other dimensions of situations or events that we talk about: At least some of the spatial locations and individuals we talk about are directly observable, but time is never directly observable but is perceived indirectly through the changes in the directly observable dimensions. We concretize time in various ways – with number lines, clocks, watches, and calendars, with conventional units such as hours (with a base of 24), minutes and seconds (with a base of 60), etc. In this dissertation, I will not delve much into the metaphysical status of time (see, e.g., McTaggart, 1908; Price, 1996), and will limit the discussion to mental representation of time as relevant to and observed through discourse production and comprehension. Nevertheless, discussion of ontology is crucial:

(1) What are times in natural language meaning: instants, intervals, or both (e.g., Kamp & Reyle, 2008)?

(2) What kinds of linguistic devices do we use to encode and preserve the structure of time in the situational content (e.g., Bittner, 2011)? How well do we make use of these devices to recover and infer the temporal structure of content?

(3) Does time in language involve simply an ordering of instants/intervals or a more fine-grained scale of magnitudes analogous to the continuous reality of our general time perception (Rinck & Bower, 2000; Jurafsky & Martin, 2008)?

(4) Do we think of time on a single linear number line stretching from the past through the present into the future, or does our mental

timeline branch into multiple possible futures (and possible pasts), equivalent to multiple timelines (e.g., Dowty, 1979; Jaszczolt, 2009)?

Answers to these questions obviously depend on how one defines *meaning*, *time in language*, etc. In subsequent chapters, I treat time in language as the mental representation of time that is relevant to the linguistic tasks of narrative production and comprehension, as expressed through explicit linguistic expressions or observed through behavioral measures such as reading speed and response times to probes. I develop a view of discourse meaning that is not tied just to the textual/verbal expressions, but encompasses broader sources of information such as perceptual access and real-world knowledge – in other words, anything that brings about changes in information states or mental representations of interlocutors engaged in discourse.[1] For question (1), I will review previous insights in theoretical linguistics and cognitive psychology in this chapter. I address the questions in (2) in Chapters 2 and 3, in which I describe my Frogbook project – an elicitation study of narratives and topic continuity judgments with wordless picture books. I address question (3) in Chapter 4, in which I describe my Two-Minute Mystery project – a study of the effect of duration in temporal adverbial phrases on reading times and word recognition latencies.

1.1.1. The Language of Time

In early research on linguistic expressions of time, Litteral (1972) proposed a topological account of time in language, de-emphasizing the exact duration of an event and emphasizing ordering of events instead. Grammatical markers of time in many of the

[1] I am not limiting discourse to synchronous conversation. In fact, most of the studies I review and report involve offline communication – monologue narrative to an imaginary audience, prepared (though naturalistic) stories for participants to read, etc.

better-known languages which indicate only the temporal direction of an eventuality relative to the speech time or the presence/absence of temporal overlap between eventualities reinforce such a topological view. In these languages, expressing more fine-grained temporal magnitudes requires explicit lexical or phrasal adverbials, such as *immediately*, *after three minutes*, *for two hours*, etc.

Dahl (1984) and Comrie (1985), however, reviewed 'metrical tense' systems in languages that make distinctions between degrees of temporal distance or remoteness, e.g., immediate past vs. recent past vs. distant past. Dahl (1985) thus concluded that "objective time measures […] play an important role in determining the choice between different tenses" (p. 123) in these languages. In a more recent study, Bohnemeyer (2009) described aspect-mood markers in a tenseless language, Yucatec Maya, as 'metrical predicates' that "cardinally quantify over the distance between topic time and event time" (p. 95) and are dependent on contextual standards for judgment of proximity/remoteness. In short, the availability of lexical and grammatical means for expressions of temporal remoteness or magnitude raises interesting questions about the granularity of temporal representation that supports discourse across languages.

1.1.2. Time in Formal Semantics

1.1.2.1. Ontology: Instants, intervals, and scales.

Our most common, intuitive conceptualization of metaphysical time is analogous to the strict linear ordering of the real numbers, as in classical physics (Kamp, 1979; Bittner, 1999). For time in language, however, early insights have pointed to necessary discretization of the continuous reality. Kamp (1979) argued that, although there are no truly durationless events in the physical world, some events can "play the role of instants"

(p. 404; see the notion of ‘contraction’ on p. 405) in discourse representation. In addition, the very notion of temporal succession in the expression *and next* requires a discrete scale, although the adjustment is subjective and gives rise to occasional disagreements between interlocutors regarding what was the appropriate ‘next instant.’ Bittner (1999) presented a two-pronged argument for a discrete structure of time in language. First, “no language has temporal expressions that require continuous temporal order, whereas expressions that crucially rely on discreteness are abundant [*immediately*, *throughout*, *next*, etc.]” (p. 23); and second, the event structures underlying linguistic time structures are necessarily finite, due to the “finite capacity of the brain” (p. 23). The precise determination of the size of a single point-like temporal location, or of a negligible interval (‘dead time,’ p. 25) between events that are treated as consecutive – ‘pragmatic coarse-graining’ – depends on the relevant time scale for the context.

In addition to point-like instants, intervals seem necessary in the ontology for time in language as well. Kamp (1979) noted that intervals are indispensable for truth judgment for certain eventualities: For example, whether it is true that I daydreamed about time travel a few weeks before my Ph.D. defense *in the process of* dissertation writing crucially depends on (the temporal relation of the daydreaming to) an interval unit of dissertation writing (among other factors). However, noting the problem of truth-value gaps and partiality particularly for interval-based semantics, Kamp and Reyle (2008) proposed a third alternative, an event-based account of temporal interpretation in discourse (see also Moens & Steedman, 1988, for event complexes).

In fact, it was the goal of a precise formulation of the interpretive difference between *passé simple* and *imparfait* in French discourse that motivated Kamp’s (1981)

Discourse Representation Theory. In this early account of dynamic semantics, Kamp (1981) described these markers of (viewpoint) aspect as “instructions as to the particular way in which the information that is conveyed with their help is to be ‘pictured’” (p. 17). Thus the “punctual” *passé simple* is a directive to represent a temporally ‘closed’ event and move the topic time forward, whereas the stative *imparfait* presents an ‘(event)-internal’ perspective, inside the previous event referent (see also Smith, 1991).

The following example from Kamp (1990) illustrates the representation of time in a Discourse Representation Structure (DRS; time and eventuality variables and predicates in bold):

(5) Last month a whale was beached near San Diego. Three days later it was dead.

(5’)

x **e** p z **n t t’** y **s t’’**

whale(x); **Event(e)**; Place(p); In(**e**, p); San Diego(z); Near(p, z);

Bef(e, n); Time(t); Time(t’); calendar month(t’); Succ(t’, calendar month[n]); At(e, t); Incl(t, t’)

e… Beached(x)

y = x;

State(s); Bef(s, n); Time(t’’); day[t’’] = day[t] + 3; Overlaps(s, t’’)

s… dead(y)

We can see in the DRS (5’) for the sentences in (5) that prerequisites for successful interpretation of this two-sentence discourse include knowledge of the calendrical month scale and determination of succession between months, which are examples of relevant background knowledge. The past tenses correspond to the Bef relation between the event time *e* and the utterance time *n*, with the event being before the utterance. There is also a topic time shift with the duration explicitly stated (*three days later*), which is represented in the DRS as ‘day[t’’] = day[t] + 3.’ In the framework of situation theory (Barwise & Perry, 1983), Devlin (1990) proposed an ontology including *temporal locations*, stating that we discriminate “temporal intervals of certain (but by no means all) magnitudes” (p. 81) – thus assuming both discretization and some preservation of magnitude.

As seen above, time is represented as a type of core discourse entity in most dynamic semantic formalisms, with anaphoric phenomena similar to those in other domains such as individuals and worlds (Partee, 1973; Stone, 1997; Bittner, 2011). Whether discourse interfaces with the lexical semantics of time-related words such as *day*, *month*, *year*, and number words, or indispensably involves continuous representation of time on a mental timeline even in the absence of such explicit words of magnitude, the issue of magnitude representation for time in language arises, with implications for discourse production and processing. If magnitude information in the world is preserved in linguistic representation, we need accounts of discourse representation to reflect this magnitude representation. For discourse models to be psychologically real, one may indeed expect magnitude effects for online discourse production and processing. In other words, not all topic time updates should have the same impact in real-time language use when the described durations or intervals are not the same. Differences in the numerical processing involved may lead to observable effects in linguistic performance, in the form of processing speed in comprehension or usage frequency in production. In recent theoretical literature, Fox and Hackl (2006) have made an argument for the universal density of measurement, in which our intuitions of time – “[close] to the rational or real numbers” (p. 538) – apply to all domains of measurement, including those intuitively considered to be discrete, such as individuals.

1.1.2.2. Linguistic theories and psychological reality.

Pleas for psychologically real theories of language have been made throughout the tradition of theoretical linguistics. Kamp (1981) argued in his theory of discourse representation that “representations are pictures of the world described by the sentences which determine [the representations]” (p. 2 in the original English version), and Jaszczolt (2009) similarly argued, “The way in which we represent time in semantic theory will […] have to directly reflect our conclusion on how humans represent time in thought: semantic representation will follow mental representation, as is the case in all explanatorily adequate theories of meaning” (p. 32) (see also Sanders, Spooren, & Noordman, 1992; Kehler, 2002; Hamm, Kamp, & van Lambalgen, 2006; Jackendoff, 1981; Bresnan, 1981).

Dynamic semantics has made important contributions to theory of meaning in capturing context change across sentences, but with insights from incremental psycholinguistics in the linear nature of language output and input, it is more desirable to capture incremental change in information states *within* and *across* sentences (e.g., Bittner, 2012).

1.1.3. Time in Psychological Accounts of Discourse Use

In this section and the next, I present a brief introduction to psychological accounts of discourse representation and mental representation of time. In the following chapters, I discuss in greater detail studies that are more directly relevant to my studies.

In early discussion of discourse understanding, Kintsch and van Dijk (1978) argued that the processes involved in developing a coherent representation of discourse go beyond the grammatical rules for generating the text base – a purely textual representation of discourse – and incorporate previous context, situational context, and general knowledge. van Dijk and Kintsch (1983) argued that discourse comprehension involves a 'situation model' – a model of the described situation based on an existing knowledge structure. The development of the situation model theory coincided with many other model-based accounts of language understanding (see van Dijk & Kintsch, 1983, for a review). These accounts similarly emphasized the indispensible role of knowledge and representation beyond the text – e.g., knowledge of typical situations in the form of scripts or episodes (Schank & Abelson, 1977; Anderson, Garrod, & Sanford, 1983). Schaeken, Johnson-Laird, and d'Ydewalle (1996) extended Johnson-Laird's (1983) mental-models framework to temporal reasoning with connectives *before*, *after*, and *while* in Flemish. They observed longer reading times, longer response times, and higher error rates for reasoning about "multiple-model" problems where there are multiple possible orderings (with *a happens before b* and *c happens before b*, 'a' could precede 'c' or vice versa), compared to "one-model" problems (e.g., *a happens before b* and *c happens after b*). The reading time difference, in particular, indicates rapid representation of temporal sequences, and the difference arose precisely at the premise that led to multiple models.

Time is a core dimension in most psycholinguistic accounts of discourse use, similar to dynamic semantic accounts. van Dijk and Kintsch (1983) suggested time, location, and possible world as core parameters of a situation model. Vonk, Hustinx, and Simons (1992) and Bestgen and Vonk (2000) suggested time, place, and character as three dimensions that we keep track of in discourse, and Zwaan, Langston, and Graesser (1995) argued for the five core dimensions of time, space, protagonist, causality, and intentionality in their event-indexing model. Gernsbacher (1990) also proposed time, location, causality, and reference as the core dimensions of discourse coherence in her structure-building model of language comprehension.

1.1.4. Empirical Findings

There is abundant evidence in psychological literature that temporal marking – including tense, aspect, and temporal adverbials – influences our representation of individuals or events described in language. I will limit my review to studies that involve linguistic tasks or stimuli. Moving from word/phrase-level to sentence-level (in this chapter) to discourse-level stimuli (in the following chapters), I will motivate the role of fine-grained temporal representation at all levels of linguistic tasks.

1.1.4.1. Tense and aspect.

Mental update of topic time in discourse is so prevalent that even with a sequence of simple past-tense verb forms with no explicit indication of temporal relations between eventualities across clauses, there is a default assumption of incremental progression in topic time (see, e.g., Fleischman, 1990, and Kamp, 1979, for iconicity; see Partee, 1984, Dowty, 1986, and Webber, 1988, for incremental progression based on temporal anaphora), with the relevant increments depending crucially on context.

In experimental literature, Carreiras, Carriedo, Alonso, and Fernández (1997) found that, in passages where the present tense is the dominant one, past-tense descriptions reduce the prominence of the described content in working memory, leading to longer probe recognition times. Further, the distinction between two viewpoint aspects – perfective (past perfect or simple past in English) vs. imperfective (past progressive in English) – also led to measurable differences in working memory representation in English and Cantonese speakers (Carreiras et al., 1997; Magliano & Schleich, 2000; Madden & Zwaan, 2003; Yap et al., 2009; Mozuraitis, Chambers, & Daneman, 2011). Specifically, imperfective marking led to 'ongoing' event perception and faster probe-phrase recognition times at later points in text, compared to perfective marking. Based on these findings, Magliano and Schleich (2000) argued "grammatical markers provide processing instructions for situation model construction and the maintenance of information in working memory" (p. 83), echoing the earlier intuition by Kamp (1979).

1.1.4.2. Temporal ordering.

In cognitive psychology, there has been a tradition of symbolic distance effects, starting with Moyer (1973), who found that, in a task of selecting the larger animal between a pair of animal names, the reaction time was "an inverse linear function of the logarithm of the estimated difference in animal size" (p. 180), suggesting an 'internal psychophysical judgment' in the task. In other words, larger differences in representational magnitude between stimuli lead to greater discriminability and thus less cognitive effort in discrimination tasks. In the temporal domain, evidence for rapid online representation of temporal ordering and magnitude for linguistically described events has been found in reading times, response times, and order of recall (see Franklin, Smith, &

Jonides, 2007, for a review). Franklin et al. (2007) found that deciding whether a pair of action descriptions in a routine (e.g., *Enter the Store* and *Check the Price*) were presented in the ‘correct’ (from left to right) or ‘incorrect’ (from right to left) order revealed both (a) faster responses with larger distances for unfamiliar routines, and (b) slower responses with larger distances for familiar routines. Though in opposite directions, both effects indicate that our representation of actions in a routine involves ordering and magnitudes. These effects also indicate that we make rapid use of long-term knowledge about typical sequences of actions (thus, the actions are not equidistant from one another in our representation), rather than ad-hoc comparison of two action descriptions into a simple dichotomy of ‘correctly ordered’ vs. ‘incorrectly ordered.’ The rapid impact of general knowledge has also been found for typical duration of events, with repercussions for discourse processing with no explicit discrimination task (e.g., Anderson, Garrod, & Sanford, 1983; Mozuraitis et al., 2011), as we will see in the following chapters.

Many researchers have argued for a spatial code for mental representation of time (e.g., Santiago, Lupiáñez, Pérez, & Funes, 2007; Weger & Pratt, 2008). Santiago et al. (2007) found that Spanish speakers, when presented with either ‘past’ or ‘future’ time words – including tensed verbs (e.g., *I will speak*, in Spanish), temporal adverbials (*formerly*), and temporal nouns (*future*) – were faster in deciding that a stimulus word had to do with the past when responding with the left hand, and with the future when responding with the right hand, leading to a significant interaction between temporal meaning and response location. Weger and Pratt (2008) found that this compatibility effect extended to names of actors from different time periods: In deciding whether a particular actor became popular before or after they were born, participants were faster in responding correctly to older actors with the left hand and to younger actors with the right hand, leading again to a significant interaction between time (actors' active years) and space (response location). Weger and Pratt also found a visuospatial attentional bias arising from retrospective (*yesterday, earlier, recently, past*) vs. prospective (*tomorrow, later, soon, future*) words in English: Viewing these time words before a visual categorization task – deciding whether a visual target appeared on the left or the right side – influenced the response times so that responses for the left side were faster after retrospective cues and those for the right side were faster after prospective cues, suggesting a directional mental timeline.

Beyond word-level stimuli, Ulrich and Maienborn (2010) presented whole sentences in German to investigate the effect of the same kind of time-space compatibility effect on response times in sensicality judgment tasks. The authors categorized their stimulus sentences into 'past-related' and 'future-related,' with the temporal direction established by a tense and/or a temporal adverbial. When the task was to classify the temporal direction of the sentence content as past- or future-oriented, there was a significant interaction between temporal direction and response key location on response time. Ulrich and Maienborn's results suggest, similarly to earlier studies, a 'time's arrow' for time in language and its interaction with response locations that is not based on automatic sensorimotor facilitation, but involves intermediate response coding that is improved in working memory for response selection (see Ulrich & Maienborn, 2010, for further details).

There is disagreement over the exact nature of the mental number line, with some arguing it is analogous to real-world magnitudes (Barsalou, 1999; see also, Dehaene, 2003, for a logarithmic mental number line), and others arguing it is based on discrete numerosity (Zorzi & Butterworth, 1999). If, however, there is capacity for fine-grained magnitude representation in the mind for number (Dehaene et al., 1993) and scales in natural language meaning (Fox & Hackl, 2006), there is no principled reason why representation of time in language should be limited to coarse-grained representation. Common sense tells us that full preservation of temporal magnitudes described in language is impossible – otherwise, representation of any period spanning longer than one's lifetime would last an entire lifetime and still be incomplete – and some compression is thus required. There is much psychological evidence that indicates, however, that our conceptual understanding of magnitudes preserves much fine-grained structure and continuity.

1.1.4.3. Time and space across languages: Neo-Whorfianism.

The psychological impact of the time-space mapping has also been discussed from a cross-cultural perspective, particularly with a focus on the impact of the reading/writing conventions. In early literature, Tversky, Kugelmass, and Winter (1991) found that school-age children and adults' graphic representation of temporal relations (placing stickers on paper to indicate, e.g., breakfast time and dinnertime relative to the provided lunchtime) differed across languages, as a function of the direction of the writing convention. English speakers (with left-to-right writing) showed a strong preference for left-to-right representation (from earlier to later in time), Arabic speakers (with right-to-left writing) showed a strong preference for right-to-left representation, and Hebrew speakers (with right-to-left writing, but with weaker directionality than Arabic; see Tversky et al., p. 545) showed a weak preference for right-to-left representation, which showed a temporary developmental shift at the onset of exposure to English in formal education. Following this finding, Boroditsky and colleagues (Fuhrman & Boroditsky, 2010; Boroditsky, 2001; Casasanto, 2008) have presented the neo-Whorfian view that linguistic habits can create biases in the way we think about time, based on comparisons between English and Hebrew, and English and Mandarin (see Chen, 2007, however, for a harsh criticism from a native speaker).

Although my experiments in the following chapters are limited to English, I will discuss patterns of linguistic encoding of time that are common to human languages (e.g., explicit temporal marking and referential accessibility hierarchy), despite differences in the particular options for devices offered by different languages (e.g., frequency of pronouns, availability of zero anaphora; see also Bittner, 2012). As will be discussed in greater detail in Chapter 6, the underlying magnitude representation that supports the temporal tracking and update in discourse representation is open to nonlinguistic sources of information, and is similar to magnitude representation for other domains such as space and possible worlds.

1.2. Motivation and Outline

In the following chapters, I investigate the use and processing of explicit temporal markers and referring expressions in narrative discourse, and discuss what they reveal about the underlying temporal representation that supports the discourse. In Chapters 2, 3, and 4, I report the results of a narrative elicitation study using wordless picture books. In Chapter 2, I specifically discuss the relationship between intervals in story time and the use of explicit lexical and phrasal temporal markers. In Chapter 3, I discuss the relationship between intervals in story time and referring expressions for referring back to characters introduced earlier in the narrative. I hypothesized specifically that the longer the inter-event interval in the story, (1) the more explicit the temporal marking as a device for temporal update in discourse representation, and (2) the more frequent the use of proper names rather than anaphoric expressions due to a greater discontinuity in the situation model of discourse. In Chapter 4, I explore the utility and limitations of some currently available tools in computational linguistics for automatic analysis of referring expressions, with suggestions for improvement.

In Chapter 5, I discuss the Two-Minute Mystery experiment, in which I extended the earlier insights from studies reporting a ‘narrative time shift’ or ‘temporal distance effect’ (Zwaan, 1996; Rinck & Bower, 2000; Kelter, Kaup, & Claus, 2004; Speer & Zacks, 2005). Designs of earlier studies of narrative processing have been limited to either only two levels of narrative time interval or three levels with different units of time (e.g., *a moment later* / *an hour later* / *a day later*), so I tested three levels of narrative time shift in the same unit of minutes (*10 minutes later* / *20 minutes later* / *40 minutes later*). I hypothesized that reading times and probe word recognition latencies would increase with each higher level of temporal duration.

My experiments in the following chapters will provide insight into the extent of the mental representation of temporal magnitudes in situational content in both discourse production and comprehension.

Chapter 2

The Frogbook Project: Temporal Marking in Discourse

2.1. Introduction

2.1.1. Overview

There are a wide variety of syntactic categories of temporal expressions. In English, connectives or conjunctions such as *while, since, until,* and *after* (e.g., *after the boy went to bed*), adverbs such as *then, immediately,* and *suddenly*, prepositions such as *during*, *before*, and *after* (e.g., *after an exam*), and tense and aspectual marking on verbs indicate the temporal relationship between eventualities within a sentence or across sentences. In storytelling, these expressions are typically used to relate a pair of eventualities that are both remote from the utterance time or situation (recounting past or imaginary events and states of affairs). Their use in storytelling can thus reveal how we mentally represent the temporal dynamics in discourse when not relying much on the physical environment.

When we tell a story, we don't always describe a continuous flow of events, but often introduce shifts in story time. Our use of temporal markers in narrative discourse can thus provide insight into one of the ways in which situational dimensions of target content are preserved in our linguistic coding. In this chapter, I will first review earlier studies employing discourse analysis as well as passage continuation and reading comprehension experiments. I will then describe our narrative elicitation project based on wordless picture books, along with our results on temporal marking in narratives. I will focus on lexical and phrasal means for explicit temporal marking, such as connectives, adverbs, and phrasal adverbials. Because lexical and phrasal markers are optional unlike grammatical markers such as tense in English, they can be analyzed for conditional frequencies.[2] In other words, we can ask questions such as, Are lexical and phrasal markers more frequent under certain conditions than others (e.g., after a temporal shift in the story)?

2.1.2. The Use of Temporal Markers and Its Impact on Comprehension

In their analysis of naturally occurring discourse, Schiffrin (1987) and Tenbrink (2007) emphasized the temporal continuity indicated by *then*. They observed that clause-initial or mid-clause *then* is used to describe sequences of events in immediate succession, whereas clause-final *then* is used to refer back to a previously specified time. Asher and Lascarides (2003) argued *then* also marks thematic continuity, whereas Gernsbacher (1990) and Bestgen and Vonk (2000) categorized *then* (along with time-anchoring adverbials such as *around eleven o'clock*) as 'adverbial leads' or 'segmentation markers.'

Phrasal adverbials have also been studied in reading experiments that test text-mapping vs. model-mapping accounts of discourse comprehension. Anderson, Garrod, and Sanford (1983) found an effect of the duration of story time gaps (indicated by temporal adverbials) on the frequency and type of character re-mentions in their participants' passage continuation. Specifically, story time gaps that went beyond the typical duration of the target episode (e.g., *seven hours later* in a scenario of 'going to the movies') reduced the overall mention of secondary characters, compared to time gaps within the typical duration (*ten minutes later*). Similarly, Zwaan (1996) found that

[2] Particularly in our narratives, verb tenses were not only obligatory, but also showed little variation. Participants maintained a dominant tense, typically the past tense, throughout a narrative with few tense shifts.

temporal adverbials indicating temporal discontinuity (*an hour later* and *a day later*) led to both longer reading times and longer response latencies in probe recognition[3] compared to an adverbial indicating temporal proximity (*a moment later*). These authors argued that these results support a situation model of discourse comprehension, rather than a text-mapping account.

Bestgen and Vonk (2000), who studied French temporal adverbials indicating a temporal location (e.g. *around eleven o'clock*) rather than gap duration (e.g., *an hour later*), found that temporal-location adverbials eliminated the so-called 'boundary effect' (Haberlandt, Berian & Sandson, 1980): Although participants did take longer to read topically discontinuous transitions in narrative discourse (*I dressed myself warmly. => I cut up a slice of cooked ham.*) than topically continuous transitions (*I put the roast in a saucepan. => I cut up a slice of cooked ham.*), when discontinuous transitions were preceded by a temporal marker (e.g., *around eleven o'clock*), there was no reading-time difference between continuous vs. discontinuous transitions. It wasn't just any kind of adverbial that eliminated the boundary effect: There was a significant reading time difference between topic continuity and discontinuity conditions for another kind of adverbial, *as usual*, which describes neither a topic time nor a time shift. Bestgen and Vonk (2000) thus argued that temporal markers serve as segmentation markers in discourse, triggering a new partition or structure in the reader's discourse representation and bypassing an attempt to integrate new information with the current representation.

[3] The probe recognition paradigm (McKoon & Ratcliff, 1980) is widely used in reading experiments as a measure of accessibility of prior information in working memory. In this paradigm, the participant decides whether a probe word or phrase appeared in earlier discourse. The assumption is that the less accessible a target word is in working memory, the longer the participant takes in recognizing it.

In sum, corpus studies and production and comprehension experiments demonstrate that various kinds of temporal expressions mark temporal and discourse continuity/discontinuity, and rapidly impact the speed of processing and planning of subsequent narrative. Locating the topic time of discourse at a specific point in time triggers an update in discourse representation in a manner that reduces surprisal at an unexpected shift in content. Further, the duration of an interval denoted by a temporal adverbial affects how fast we process text, and how accessible information prior to the temporal adverbial is in our working memory (see Chapter 4). I will argue below that the exact discourse function of a marker depends on its grammatical category (e.g., connective, adverb, etc.) and sometimes even on the lexical meaning of the particular word (e.g., *then, immediately, finally*).

2.1.3. Dimensions of Situational Content

In an account of situational representation of discourse, an important issue is what kinds of things we keep track of. Global coherence has been commonly called thematic or topic continuity[4] (Vonk, Hustinx, & Simons, 1992; Magliano, Zwaan, & Graesser, 1999), but it consists of multiple sub-dimensions. As discussed in Chapter 1, most accounts of discourse representation (e.g., Grimes, 1975; Gernsbacher, 1990; Vonk et al., 1992; Zwaan, Langston, & Graesser, 1995; Bestgen & Vonk, 2000) have treated time, space, and protagonist as core dimensions. In fact, Magliano et al. (1999) found evidence from sentence fit judgments (asking how well sentences fit into the story context) and reading times that these sub-dimensions affect our perception of story continuity differently. In our study, we pitted time against the global dimension of topic

[4] It is important not to confuse this with 'topic' or 'topic continuity' in the technical sense in linguistics (e.g., in the context of Centering; Brennan, 1995).

by obtaining scene-by-scene topic continuity estimates (see Section 2.2.3). We will argue that the adverb *then* is a lexical item that dissociates the temporal dimension from the global topic dimension, because it indicates immediate succession and continuity in time, though possibly a discontinuity in topic.

2.1.4. Motivation for Our Stimuli and Task

We investigated the relationship between passage of time between events in a story and the frequency and type of temporal markers (excluding grammatical markers such as tense and aspect) in narrative production. Narratives typically have a globally sequential and coherent storyline and focused target content, and they are thus particularly well-suited for a study of the dynamics of discourse production and comprehension.

We presented wordless picture books by Mercer and Marianna Mayer (Mayer, 1969; Mayer, 1974; Mayer & Mayer, 1975) to elicit narratives. These picture books provide fixed and clear target content in a coherent storyline, with no verbal intervention. These books are thus well-suited for studying the impact of nonlinguistic situational dimensions in content on narrative production. Moreover, because the time intervals between events in consecutive scenes vary and there are multiple characters engaging in different actions simultaneously in each scene, they provide ample opportunity for the use of temporal and referring expressions in narrative production (see Chapter 3 for referring expressions).

Although ideally we would manipulate each situational dimension in turn while holding the other constant, in practice it is difficult to do so. For example, it would require telekinesis to effect a spatial shift while maintaining temporal and character continuity. Fortunately, manipulating temporal discontinuity while holding other dimensions is possible, and easier particularly when there are visual cues (e.g., the sun, the moon, or a clock). Indeed, some scenes in our picture books contained such visual cues, providing valuable insight. Artificially creating too many trials of such a condition, however, would reduce the naturalness of the task or restrict the range of reasonable scenarios.

Our methodology of narrative elicitation in Chapters 2-4 is similar to earlier studies using the same stimuli (see Berman & Slobin, 1994, for a collection of such studies) and those using Chafe's (1980) "Pear Film." These earlier studies also involved visual content with no linguistic descriptions, and thus allowed the study of the relationship between situational dimensions of content and narrative production. An important difference, however, is that the participants in these earlier studies narrated orally in direct interaction with the researcher, often pausing in anticipation of approval or acknowledgement from the researcher. The elicitation of oral narrative in personal interaction provides its own unique insight that written narrative doesn't provide, but it has disadvantages as well, such as the narrators' assumption of the researcher's previous familiarity with the target content (Clancy, 1980) – which has direct impact on their referential choice – and a lack of a fit between the age of the addressee (i.e., the researcher) and the typical target age range for these stimuli (especially Mayer's picture books). The particularity of the design of earlier tasks is also evident in the adult narrators' use of the present tense as the most common dominant tense (Marchman, 1989, as reported in Berman & Slobin, 1994), against what one would expect in typical storytelling by adults.[5] Furthermore, the main focus of earlier studies using Mayer's picture books has been child narrators, rather than adults. With these picture books, it is important that the target *audience* is children, but adults can serve as natural narrators as well – storytellers for a child audience. We thus used these stimuli in our study of adults' narrative production, and removed the component of direct personal interaction in our design by presenting our stimuli and collecting narratives on the web.

2.2. Experiment 1: Estimation of Time Intervals

2.2.1. Method

2.2.1.1. Participants.

Eight native English-speaking college students (mean age = 21 years; age 20-22 years) participated. None of them reported a history of a language disorder or fluency in another language.

2.2.1.2. Materials.

We used three of Mercer Meyer's wordless picture books in the "Boy, Dog, Frog" series. These books depict the adventures of a boy and his frog, with each picture depicting a different event in the story. One book (Mayer, 1974) has 22 pictures, and the other two books (Mayer, 1969; Mayer & Mayer, 1975) have 24 pictures. In some cases, the amount of time that appears to have elapsed between pictures is quite long (see Figure 1) and in other cases it is quite short (see Figure 2).

2.2.1.3. Procedure.

[5] In our narratives, in contrast, the past tense was the most frequent dominant tense by far (28 out of 36 participants).

Participants estimated how much time had elapsed between the events depicted in each pair of consecutive pictures in the picture books. Participants wrote down their estimates, and were allowed to go back and change their time interval estimates at any point. We deliberately did not tell the participants what unit of measurement to use (seconds/minutes/hours), and participants provided the unit of measurement that they felt was appropriate for each scene transition. We used these estimates to obtain mean estimated intervals for pairs of consecutive events.

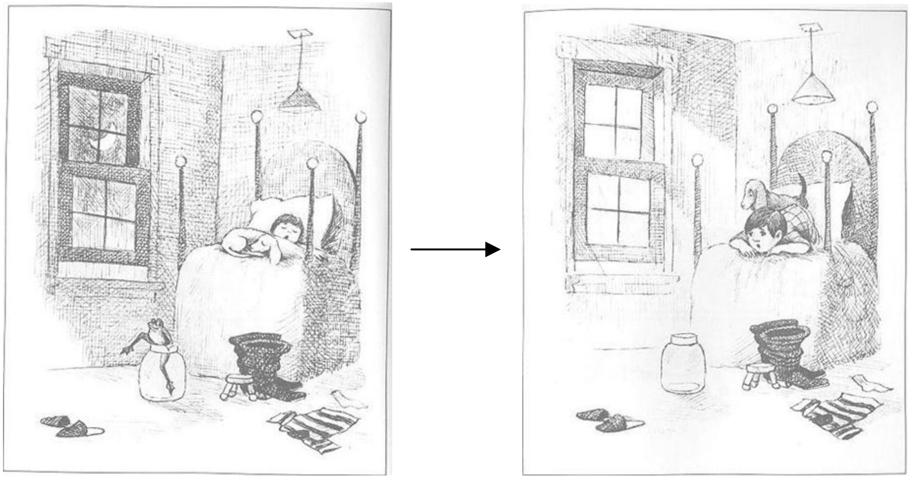

Figure 1. Sample 'Long Interval' (Mean estimate: 6h 48m 45s, between the two scenes).

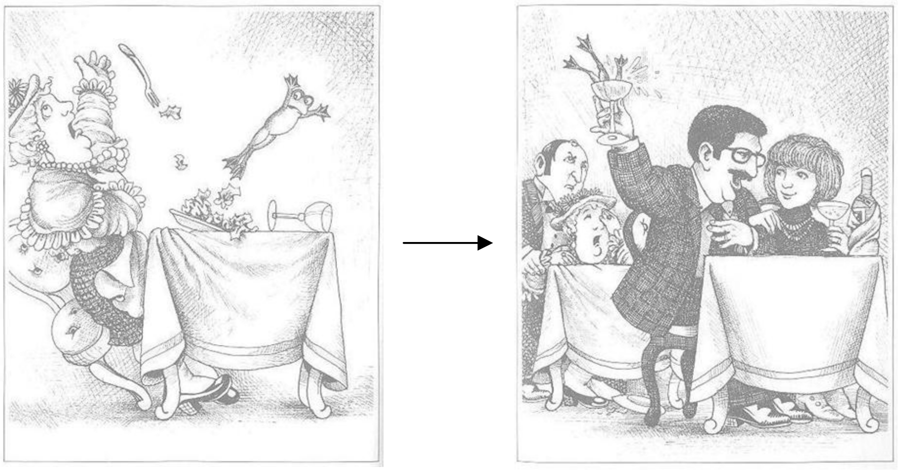

Figure 2. Sample 'Short Interval' (Mean estimate: 3s, between the two scenes).

2.2.2. Results

As we had hoped, the amount of time that was estimated to have elapsed between consecutive pictures varied considerably among the 67 scene transitions (mean = 9 minutes 26 seconds, *SD* = 49 minutes 52 seconds). The eight judges were consistent in which inter-event intervals they judged to be long vs. short: A rank-based test of inter-judge concordance showed significant agreement among the judges ($W = .30$, $\chi^2(7, N = 49) = 103.07$, $p < .001$).[6]

We used the time interval estimates from Experiment 1 to obtain mean estimated intervals for each pair of consecutive events in the three books.[7] The mean estimate for the eight longest inter-event intervals (henceforth Long Intervals) was 1 hour 7 minutes 2 seconds, and the mean estimate for the eight shortest intervals (henceforth Short Intervals) was only 10 seconds (see Table 1; see also Figures 1 and 2 for examples). In order to observe the immediate impact of an interval, we only analyzed the first sentences after the Long Intervals and the Short Intervals as our critical sentences in analyses below.

[6] Although there were 67 scene transitions in the three picture books combined, SPSS doesn't seem to take $N > 49$ for nonparametric tests of related samples. Concordances (Kendall's *W*s) for the three picture books separately were all significant, with $p < .001$.

[7] One participant wrote down '0 minutes' for a few scene pairs instead of using seconds as the unit of measurement, and we excluded these few responses from the mean calculations.

Table 1

Long and Short Intervals with Mean Estimates

Long Intervals	Mean estimated interval
Frog, Where Are You?: Scenes 2-3	6h 48m 45s
Frog, Where Are You?: Scenes 1-2	32m 45s
One Frog Too Many: Scenes 7-8	19m 23s
Frog Goes to Dinner: Scenes 3-4	18m 8s
One Frog Too Many: Scenes 18-19	17m 15s
Frog Goes to Dinner: Scenes 20-21	14m 45s
One Frog Too Many: Scenes 19-20	13m 8s
Frog Goes to Dinner: Scenes 4-5	12m 8s
Short Intervals	Mean estimated interval
Frog Goes to Dinner: Scenes 13-14	3s
One Frog Too Many: Scenes 14-15	6s
Frog, Where Are You?: Scenes 17-18	7s
Frog Goes to Dinner: Scenes 12-13	7s
Frog, Where Are You?: Scenes 18-19	9s
One Frog Too Many: Scenes 22-23	11s
One Frog Too Many: Scenes 2-3	15s
Frog Goes to Dinner: Scenes 8-9	20s

2.3. Experiment 2: Story-writing

We hypothesized that the longer the interval between depicted events, the higher the frequency of temporal markers that update the topic time with a temporal specification.

2.3.1. Method

2.3.1.1. Participants.

A different group of 36 native English-speaking college students (mean age = 20 years; age 18-22 years) wrote stories to accompany Mayer's books. None of them reported a history of a language disorder, and two were late learners of a foreign language.

2.3.1.2. Materials and procedure.

Participants were told to write stories that children would listen to while looking at the same picture books as in Experiment 1. The first 11 participants wrote a story for each of the three books, but a few of them took over two hours to write all three stories. Thus, we had the next 25 participants write a story for only one of the books (*Frog Goes to Dinner*, Mayer, 1974).[8] Each participant was assigned one of the following three Conditions:

(a) "Planning & Editing" Condition (a total of 23 narratives): Participants looked through the pictures first before writing a story, and could revise their story as they saw fit;

(b) "Editing" Condition (a total of 15 narratives): Participants did not look through the pictures first before writing, but could revise their story;

[8] There were a couple of other modifications for the group of 25 participants: In order to encourage detailed narrative explicitly, we included a model page (with four sentences and 61 words) from a picture book with text (Holmes, 1977), and modified the target age of the imaginary child audience from six to eight years.

(c) “No Editing” Condition (a total of 20 narratives): Participants did not look through the pictures first, and could not revise their story.

Pictures were presented one at a time on a computer screen. The picture sizes were either 400 pixels wide and 500 pixels long (for scenes that were on one page), or 640 pixels wide and 400 pixels long (for scenes that were on two adjacent pages). To the left of the picture, a vertical thumbnail gallery with a scrollbar (15% of the width of the screen) provided thumbnail views of the entire sequence of pictures in a picture book. For each picture, participants typed the text in a text box (size: 10 rows, and 100 characters in a row) that appeared beneath the picture. They were not given any instructions on what or how much to write. We conducted pre-tests to ensure compatibility on various environments, and no incompatibility issues were reported by the participants.

2.3.2. Results

2.3.2.1. Text length.

We elicited a total of 58 written narratives, which contained a total of 2,778 sentences and 38,936 word tokens. The average narrative had 48 sentences and 671 word tokens. We analyzed various text length measures to investigate the relationship between (temporal) Interval and text length, based on the first sentence after each of the critical eight Long Intervals and eight Short Intervals as a critical sentence. In the Long (temporal) Interval category, there were a total of 159 critical sentences and 2250 words, and in the Short Interval category, there were a total of 163 critical sentences and 1941 words. For the narratives with self-revision (in the Planning & Editing and the Editing Instructions), only the final versions were analyzed. An independent-samples *t*-test indicated that participants wrote more after Long Intervals than after Short Intervals (mean length = 14.15 words and 11.91 words, respectively; $t(320) = 3.63$, $p < .001$, two-tailed). This was due in large part to the higher frequency of lexical and phrasal temporal markers (e.g., *when …*, *as soon as …*) after Long Intervals compared to Short Intervals.

2.3.2.2. Temporal connectives and adverbials.

We predicted that temporal marking with Temporal Connectives (e.g., *when, while, since, as*), Temporal Adverbs (e.g., *then, finally, now*), or Phrasal Time References (e.g., *the next day, in the morning, during the night*) would be more frequent in sentences after Long Intervals than after Short Intervals. Approximately 8% of the critical sentences (27 out of 322) were excluded because they explicitly marked for a non-temporal dimension, such as space (e.g., *Inside*), causation (e.g., *because*), or rhetorical relation (e.g., *Although*). As predicted, explicit temporal marking was significantly more common after Long Intervals (86 out of 149 scenes) than after Short Intervals (43 out of 145 scenes) ($\chi^2(1, n = 294) = 27.01$, $p < .001$, $\varphi = .30$).[9] (See examples in (1) and (2).)

(1) *Once* Eric was in bed, Bob decided to sneak out of his jar.

(Long Interval, 'Marking')

(2) Frankie landed in the water with a loud plop.

(Short Interval, 'No Marking')

[9] We conducted a chi-square test on raw counts to avoid doing an ANOVA on proportions (see Jaeger, 2008). Some would argue that the chi-square analysis we performed is also problematic because multiple observations from each participant were combined into the totals. We thus conducted a mixed-effects logistic regression analysis as well.

A mixed-effects logistic regression analysis with crossed random effects of Participant and Story (Quené & van den Bergh, 2008) confirmed that the Interval effect was significant ($t = 5.06$, $N = 294$, $p < .001$).[10]

Frequency patterns by Interval differed among types of temporal marking: Temporal Connectives and Explicit Time References were more common after Long Intervals than after Short Intervals, whereas Temporal Adverbs were more balanced ($\chi^2(2, n = 139) = 6.62$, $p < .05$, $\varphi = .22$; see Figure 3, 'TC' for Temporal Connectives, 'TA' for Temporal Adverbs, 'Other' for Explicit Time References, 'LI' for Long Intervals, and 'SI' for Short Intervals). Compared to Temporal Connectives, Temporal Adverbs make more heterogeneous contributions to discourse. For example, *then* was used 11 times after Short Intervals to mark immediate succession, but never after Long Intervals, whereas *finally* occurred only once after Short Intervals and four times after Long Intervals.

[10] According to Baayen (2008), there is no standard way yet of determining the exact *df* and thus the significance for a the *t*-value in linear mixed-effects modeling, but a heuristic for large *N*s is to use the *N* for the look-up table of critical values.

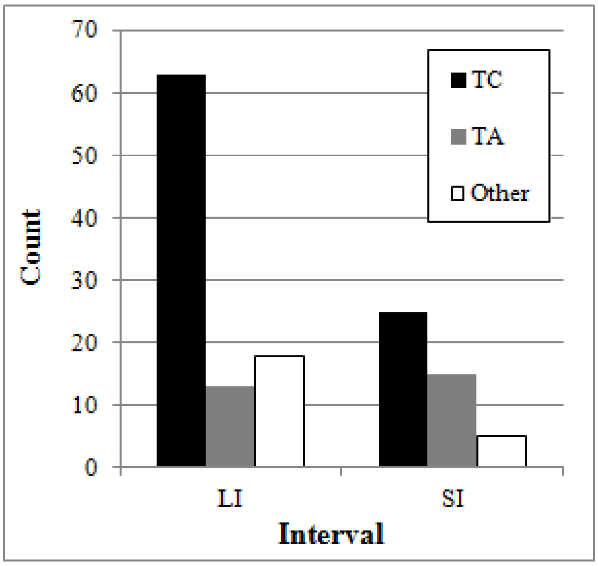

Figure 3. Frequency counts of temporal marker types by Interval.

Linear mixed-effects modeling with Participant as a random factor indicated that the effect of (Instructions) Condition on frequency of explicit temporal marking was non-significant ($t(296) = -1.45$, $p = .15$, two-tailed).[11]

2.4. Experiment 3: Topic Continuity Judgments

In order to pit temporal continuity against topic continuity as predictors of discourse production, we obtained estimates of topic continuity.

2.4.1. Method

A different group of 12 native English-speaking college students (mean age = 18.5 years; age 18-19 years) participated.[12] Scenes from the three picture books were

[11] In our previous results (Lee, Kharkwal, & Stromswold, 2012), Welch's *t*-tests indicated some effect of Instructions Conditions on frequency of temporal marking, with more explicit temporal marking for the "Planning & Editing" Condition over the others. The effect, however, was not robust.

[12] None of the participants reported a history of a language disorder, and two participants were bilingual.

presented sequentially, two at a time. Participants judged “whether the topic of the story changes from the old scene to the new scene.”

2.4.2. Results

Proportions of topic change judgments from 12 new judges varied considerably among scenes, ranging from 0% for some scenes to 85% for other scenes (mean = 26.18%, *SD* = 17.87%). For each of the 67 scene transitions, we used the data obtained in Experiment 1 to calculate the mean inter-scene time interval and the data obtained in Experiment 2 to calculate the mean topic change proportion. The correlation between the mean time interval and the mean topic change was not significant ($r = .034$, $N = 67$, $p = .78$), but the rank orders for the two sets of estimates were significantly correlated ($r_s = .46$, $N = 67$, $p < .001$). In other words, temporal continuity is distinguishable from general topic continuity, but it seems to be a component that contributes to topic continuity, consistent with earlier insights (Vonk et al., 1992). Regression analyses pitting topic continuity against temporal continuity as predictors of temporal marking indicated that time interval estimates were a slightly better predictor of temporal marking ($R^2 = .039$, $b = .197$, $t(294) = 3.52$, $p < .001$) than topic change proportions were ($R^2 = .035$, $b = .188$, $t(294) = 3.35$, $p = .001$).

2.5. Discussion

Narrators use linguistic temporal expressions such as Temporal Connectives and Explicit Time References to update the topic time in discourse. The observed relationship between perceived interval between events in stories and the frequency and type of temporal transitions in narratives demonstrates the impact of story time on narrative organization, and is consistent with situation models.

Explicit temporal references such as *after a long 30 minute drive* and *on the first day of school* are a straightforward means of updating the topic time in discourse. Temporal Connectives such as *when* and *as soon as* in subordinate clauses similarly update the topic time by specifying the time of the event in the main clause. These updates in discourse representation may facilitate the reader's discourse processing by triggering a new partition in the mental representation of the discourse (Bestgen & Vonk, 2000).

Temporal Adverbs, on the other hand, play a more heterogeneous role in narrative time dynamics, as diverse lexical meanings in this category seem to have direct consequences for their discourse role of relating a new topic time to the previous topic time (in this case, describing the temporal relationship between scenes across an interval). For example, *then* (the only temporal marker that appeared noticeably more often after Short Intervals in our data) typically indicated immediate succession or continuity in time, whereas *finally* appeared once after Short Intervals and four times after Long Intervals. Our data thus suggest that *then* plays a role in marking temporal continuity in the described situations (Schiffrin, 1987; Tenbrink, 2007), thus reinforcing the default incremental narrative time progression (Dowty, 1986). Further investigation is required to determine whether *then* maintains the current unit of discourse representation with temporal and topic continuity (Asher & Lascarides, 2003), or triggers a new unit of discourse representation due to a topic shift (Gernsbacher, 1990; Bestgen & Vonk, 2000).

Because of their meanings we expected *immediately / at once* to be used more after Short Intervals than after Long Intervals, but all four occurrences were after a Long Interval. Closer examination revealed that all of these cases involved multi-clause sentences in which there was an earlier time introduced by another temporal marker, and *immediately* described immediate succession between two events within the same sentence (thus with a discourse function similar to *as soon as*), rather than between two events across a scene transition:

(3) The minute they got home[,] Zach's dad told Zach to go to his room *immediately*.

In addition, our participants wrote longer descriptions for the scenes after Long Intervals than after Short Intervals, and the first sentences in these critical scenes were longer after Long Intervals than after Short Intervals. This is expected in light of the fact that temporal markers such as phrasal adverbials and Temporal Connectives leading a subordinate clause were more commonly used after Long Intervals.

For the addressee, explicit temporal markers seem to serve in general as cues for discontinuity in story time, based on our preliminary results from a follow-up study. In this segmentation experiment, a different group of participants marked where they thought scene boundaries were in three sample narratives written by one of our participants in Experiment 2 (see Section 2.3), while looking at the corresponding picture book for each narrative. (*Then* never appeared in the three narratives, so the temporal markers in these narratives were mostly markers of a discourse break.) Participants' detection of scene boundaries was significantly more accurate for scene transitions that were temporally marked compared to those that weren't (mean = 89.1% vs. 79.7%; $\chi^2(1, N = 448) = 6.11$, $p = .014$; $t(446) = 2.48$, $p = .013$, two-tailed).

Effects of nonlinguistically presented interval on explicit temporal marking in narratives reflect the pervasive impact of the goal of alignment of situation models in discourse (Pickering & Garrod, 2006), even when the participants were told that their reader would see the picture books as their stories are read to her/him. The relevance of model-based accounts (Johnson-Laird, 1983; Garnham, 1991; Zwaan, 1999) extends to discourse production and thus both sides of discourse – a cooperative effort to align situation models.

Chapter 3

The Frogbook Project: Referring Expressions and Accessibility

In addition to the topic time of discourse, discussed in the previous chapter, characters are another kind of information a narrator and the addressee must keep track of in narrative discourse. The narrator can use a variety of referring expressions to re-mention characters that have been introduced in earlier discourse. For example, she can repeat a name (e.g., *Billy*) or use anaphoric devices, such as a role label with a definite description (*the student*) or a pronoun (*he/him*). These referential types have different degrees of specificity or identificational explicitness (Vonk, Hustinx, & Simons, 1992), with proper names being the most explicit and pronouns having the least descriptive content. Referring expressions are often more specific than necessary for referent identification and coreference resolution (Pechmann, 1989). For example, narrators often repeat a full definite description or a proper name even when there are enough cues that a pronoun would suffice to identify the referent uniquely. This suggests that referring expressions may serve an additional discourse function. In this chapter, I will review earlier work relating types of referring expressions in discourse to information structure or accessibility of referents, and analyze the use of referring expressions in our Frogbook narratives in relation to inter-event intervals. I argue that the narrator's referential choices are another way of indicating the temporal structure of the situational content.

3.1. Introduction

3.1.1. Information Structure and Accessibility/Salience

Scholars in various fields have studied the relationship between the speaker's referential choice and discourse (dis)continuity and noted the role of referential type as a marker of information structure (e.g., Lewis, 1979; Prince, 1981). The type of referring expression a narrator uses in mentioning a discourse entity reflects whether the referent is 'given' or 'new' in the discourse (see Prince, 1981, for a review of early theoretical work on this distinction). Chafe (1976), Prince (1981), and others made a connection between the given-new distinction and the addressee's 'consciousness' or mental representation, thereby putting discourse processes in the context of broader cognitive factors of accessibility in working memory. Others have proposed specific accessibility orderings for different referential types based on analysis of naturally occurring discourse (Ariel, 1990; Givón, 1992; Gundel, Hedberg, & Zacharski, 1993). The consensus ordering is depicted in Figure 4:

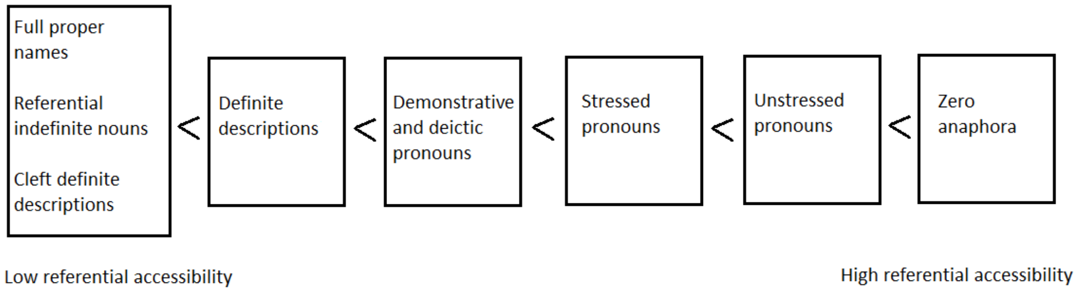
Full proper names
Referential indefinite nouns
Cleft definite descriptions
<
Definite descriptions
<
Demonstrative and deictic pronouns
<
Stressed pronouns
<
Unstressed pronouns
<
Zero anaphora
Low referential accessibility
High referential accessibility

Figure 4. Referential type ordering based on accessibility of referents.

While the general ordering of referential types may hold quite robustly across languages, the exact point in the ordering at which referential types shift from marking *continuity* to marking *discontinuity* differs from language to language, based on the availability or typicality of the referential type options in the spectrum (see Clancy, 1980, for a comparison between English and Japanese). For example, in English, where zero anaphora is typically not an option, pronouns are typical *continuity* markers, toward the anaphoric end of the spectrum. In East Asian languages, where missing arguments are much more common and pronouns are very infrequent compared to English, missing arguments play the role of *continuity* marking, and the infrequent pronouns, when they do appear, often indicate a shift or break in discourse (e.g., a change in topic; Li, 2004, for Chinese). In this chapter, I will restrict my discussion to proper names, definite descriptions, and personal pronouns in general, as other categories were scarce in our data.

Factors such as the textual distance between the antecedent and the re-mention (e.g., Givón, 1992), the presence of a referential competitor (e.g., Arnold & Griffin, 2007), topicality of the antecedent (e.g., Brennan, 1995), and thematic or episodic structure (Clancy, 1980; Anderson, Garrod, & Sanford, 1983) affect the accessibility of referents and, thus, referential choice. While most of these factors have a *textual* basis, it is generally agreed that, in addition to *linguistic* context, the speech situation or physical environment and shared general knowledge can also affect referential choice. For example, episodic structure requires integration of knowledge about typical situations (Anderson et al., 1983).

In addition, Arnold and Griffin (2007) found that, even in cases where a pronoun wouldn't have been referentially ambiguous, presence of a second character in an earlier scene reduced pronoun use. This Two-Character Effect was present regardless of whether the second character was explicitly mentioned in the opening statement or not. In other words, it was the visual presence of another entity in the (earlier) scene or situational context, rather than the linguistic presence of an antecedent in discourse context, that gave rise to the competition effect – a point often missed in earlier literature. Recent studies by Fukumura, van Gompel, and Pickering (2010) and Kantola and van Gompel (2011) also demonstrated referential interference from a visually available entity, indicating that nonlinguistic context affects referential choice. The question of referential accessibility is essentially the question of the nature of discourse representation: What kinds of entities are represented in our mental representation of discourse? Is discourse representation purely textual, involving just the hierarchical structure and meaning of the linguistic expressions, or does it also involve representation of nonlinguistic content?

3.1.2. Situation Model of Discourse

The broadening perspective in the debate on information structure coincided with van Dijk and Kintsch's (1983) emphasis on the situation model – a level of representation beyond just the surface form of text. van Dijk and Kintsch argued that discourse representation should include not just the surface form of the discourse but also what the discourse is about. Psycholinguistic work on the situation model of language meaning has focused on comprehension (e.g., Zwaan, 1999). We believe this has been the case for two main reasons. First, because the listener often has less access to the target situational content than the speaker, it is the listener who must make inferences about the situation based on linguistic input. Secondly, comprehension studies readily lend themselves to well-controlled experiments for standard behavioral measures such as reading times and decision latencies, whereas production data are open-ended and unpredictable. The resurgence of research activity surrounding the situation model of language meaning (e.g., Zwaan, 1999), communication as bidirectional alignment of situation models (Pickering & Garrod, 2006), and audience design (Clark & Murphy, 1982; Barr & Gann,

2011) calls for a closer look at the relationship between situational dimensions and language production.

3.1.3. Referring Expressions in Discourse Production and Comprehension

In a passage continuation experiment, Sanford, Moar, and Garrod (1988) demonstrated discourse functions of referring expressions beyond referent identification by showing that participants mentioned characters that were introduced with a proper name more frequently in their continuations compared to those introduced with a definite description. In addition, participants were faster at reading pronominal anaphors with a proper-name antecedent than those with a definite-description antecedent.

In a Dutch passage continuation study on the relationship between situational continuity and referential type, Vonk et al. (1992) found that when the feeder word provided for the protagonist was a pronoun, participants wrote thematically *continuous* sentences more often than *discontinuous* ones. Full NP feeders (names and definite descriptions) led to the opposite pattern. Further, when participants read Dutch narratives elicited with wordless comic strips, pronouns led to perception of a thematic *continuation*, and full NPs led to perception of a thematic *shift*. In self-paced reading experiments with probe recognition tasks, Vonk et al. (1992) also found that, compared to a pronominal re-mention of the protagonist, a full NP re-mention led to slower recognition of a probe word from the sentence preceding the re-mention, possibly due to the creation of a new representational unit at the full NP. In sum, their results suggest that referential type affects both language production and language comprehension.

In a passage continuation experiment testing model-mapping vs. text-mapping accounts, Anderson et al. (1983) focused on discontinuity in the narrative timeline as a specific dimension of global continuity in situational content. They found that long temporal intervals indicated by adverbial phrases (e.g., *Five hours later*, as opposed to *Forty minutes later*) led participants to produce a lower proportion of pronominal anaphors referring to secondary characters compared to protagonists. Anderson et al. concluded that referential choice is sensitive not only to textual factors but also to content-based factors of accessibility, such as duration of a shift in story time.

3.1.4. Motivation and Hypothesis

In the current study, I investigated whether nonlinguistically presented situational dimensions of the storyline serve as additional factors of salience in discourse organization. I hypothesized that the longer the interval in story time between two events, the more explicit the referring expression used for re-mentioning a character. Participants, materials, and tasks were identical to those described in Chapter 2. In the next section, I describe the coding schemes and results for referring expressions.

3.2. Results

We analyzed referring expressions in the critical sentences (the first sentence after each of the eight Long and the eight Short Intervals; see Chapter 2). We excluded direct quotes and referring expressions that introduced a character for the first time or referred to an inanimate object. To avoid long textual distances between the antecedent and the re-mention, we excluded referring expressions that referred back to a character that was not mentioned in the immediately preceding scene. To avoid intra-sentential anaphora, we only counted the first mention of each character in the critical sentence.

We manually coded each referring expression as a Proper Name (e.g., *Mr. Frog*), a Definite Description (e.g., *the frog*), or a Pronoun (e.g., *he*). Pronouns included quantifiers such as *everyone* and possessive forms such as *her* and *his*. Each referring expression was coded by two people, and disagreements were resolved via discussion. The frequencies of Referential Types by Interval are presented in Table 2.

Table 2

Raw Counts and Mean Numbers of Occurrences (and SDs) of Referring Expressions by Interval and Referential Type

		Long Interval		Short Interval		Overall	
		Count	Mean	Count	Mean	Count	Mean
Proper Name		120	0.72 (0.68)	67	0.39 (0.61)	187	0.55 (0.66)
Definite Description		53	0.32 (0.54)	74	0.43 (0.57)	127	0.38 (0.56)
Pronoun	Singular	6	0.036 (0.49)	30	0.18 (0.43)	36	0.11 (0.46)
	Plural	44	0.27 (0.44)	7	0.041 (0.20)	51	0.15 (0.36)
Overall		223	0.45 (0.60)	178	0.35 (0.55)	401	0.40 (0.58)

Note. Plural-denoting Pronouns were not subjected to mean-based ANOVAs.

The mean number of occurrences of referring expressions per critical sentence was analyzed via a three (Referential Type: Proper Name/Definite Description/Pronoun) by two (Interval: Long Interval/Short Interval) ANOVA with Participant as a random factor (see Figure 5: ‘PN’ for Proper Name, ‘DD’ for Definite Descriptions, ‘Pro’ for Pronouns, ‘LI’ for Long Intervals, and ‘SI’ for Short Intervals). There was a significant main effect of Referential Type (*F*(2, 789) = 12.70, $p < .001$, *MSE* = 0.19; *N* = 337 critical sentences). The two-way interaction between Referential Type and Participant was significant (*F*(72, 789) = 2.27, $p < .001$, *MSE* = 0.27), indicating that there was individual variation in the overall use of the different Referential Types. Post-hoc analyses based on Tukey’s HSD indicated that Proper Names (mean = 0.55, *SD* = 0.66) were more common than Definite Descriptions (mean = 0.38, *SD* = 0.56, $p < .001$) and than Pronouns (mean = 0.26, *SD* = 0.46, $p < .001$), and Definite Descriptions were more common than Pronouns ($p = .016$). There was also a significant main effect of Interval: Referring expressions were more common after Long Intervals than after Short Intervals (mean = 0.45 and 0.35, respectively; *F*(1, 789) = 11.76, $p = .001$, *MSE* = 0.56). The main effect of Participant was not significant (*F*(1, 789) = 0.81, $p = .73$, *MSE* = 0.52).

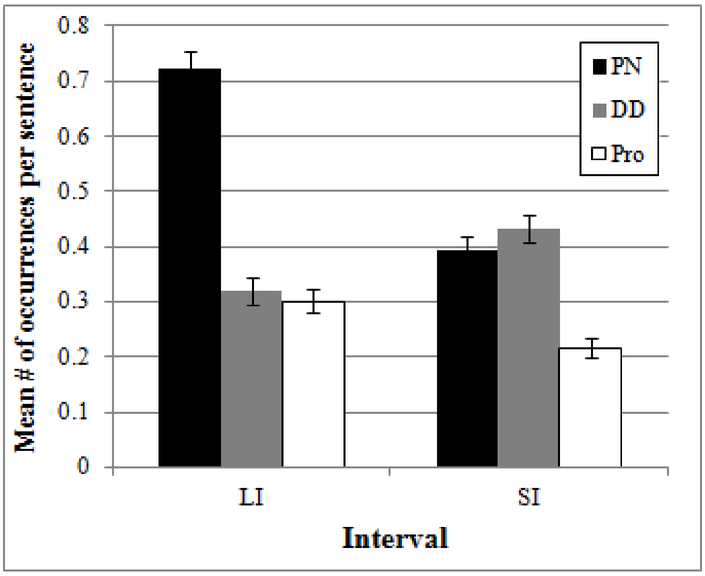

Figure 5. Mean numbers of occurrences of Referential Types per critical sentence by Interval. Error bars represent the standard errors.

We predicted that narrators would use more specific referring expressions after Long Intervals than after Short Intervals (see Figure 4), and indeed the ANOVA revealed a significant two-way interaction between Interval and Referential Type (*F*(2, 102) = 18.18, *p* < .001, *MSE* = 0.28). Planned comparisons with independent-samples *t*-tests indicated that, consistent with our prediction, Proper Names were more frequent after Long Intervals than after Short Intervals (mean = 0.72 and 0.38, respectively; *t*(320) = 4.86, *p* < .001, two-tailed); and Definite Descriptions (which are anaphoric devices) were more frequent after Short Intervals than after Long Intervals (mean = 0.44 and 0.33, respectively; *t*(320) = 1.73, *p* = .043, one-tailed). (See the next paragraph for Pronouns.) Linear mixed-effects modeling with crossed random effects for Participant and Story revealed the same pattern, with the fixed effect of Interval being significant for Proper Names (*t* = 5.09, *N* = 322, *p* < .001) and for Definite Descriptions (*t* = -1.79, *N* = 322, *p* = .037). Examples in (1) and (2) illustrate the use of a proper name after a Long Interval and the use of a definite description after a Short Interval.

(1) During the night, Earl [the frog] slipped out of his jar and escaped! /[13]
In the morning, Tom and Dog realized *Earl* was missing.
(Long Interval, Proper Name)

(2) The baby frog flew in the window! /
Everyone is so excited to have *the baby frog* back home safe and sound.
(Short Interval, Definite Description)

[13] '/' marks a boundary between two scenes, where a critical interval would be.

Contrary to our expectations, Pronouns showed a tendency toward higher frequency after Long Intervals than after Short Intervals (mean = 0.30 and 0.22, respectively; *t*(320) = 1.46, *p* = .072, one-tailed). Closer examination revealed that narrators used different types of pronouns after Long vs. Short Intervals. After Long Intervals, only 12% (six out of 50 occurrences) were singular-denoting pronouns (excluding *everyone* and *everybody*, which have plural referents), whereas after Short Intervals, over 80% (32 out of 39) were singular (see Figure 6). An independent-samples *t*-test indicated this difference was significant (*t*(87) = 9.21, *p* < .001, two-tailed).

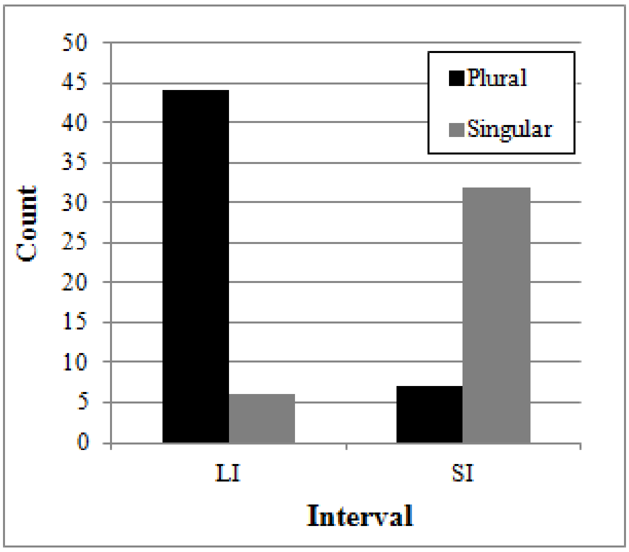

Figure 6. Frequency counts of Pronoun types by Interval.

To investigate the impact of Instructions, for each Referential Type, we performed a three (Instructions: Planning & Editing, Editing, No Editing) by two (Interval: Long Interval, Short Interval) ANOVA on the number of occurrences. Of the three Referential Types, Instructions had a significant effect only on Proper Names (*F*(2, 316) = 3.62, *p* = .028, *MSE* = 0.40; the Instructions effect for the other two Referential Types were non-significant, with both *F*s < 1.94). Post-hoc analyses revealed that narrators who were allowed to plan and edit their stories used more Proper Names than narrators who were allowed to edit – but not plan – their stories (mean = 0.63 and 0.38, respectively, p = .018; other pairwise comparisons were all non-significant, with *p*s > .10).

We also analyzed other factors that have been found in earlier literature to affect the salience of referents in discourse (Givón, 1992; Brennan, 1995; Arnold & Griffin, 2007). First, Textual Distance – defined as the number of intervening words between the antecedent and the re-mention – for Pronouns (mean = 7.02 words) was significantly shorter than that for Proper Names (mean = 12.73 words; Tukey's HSD, p < 0.001) and that for Definite Descriptions (mean = 13.25 words; Tukey's HSD, p < 0.001). Textual Distance thus seems to have a partial role in the choice between a pronoun vs. a nominal re-mention, but it cannot account for the Referential Type patterns across Interval conditions in our results (such as the high frequency of Proper Names after Long Intervals). Second, Antecedent Subjecthood – the proportion at which a referring expression in a critical sentence was coreferential with the *subject* of the preceding clause (Brennan, 1995) – was highest for Pronouns (mean = 55.0%), as a Centering account would predict; however, this proportion for Pronouns was significantly higher only compared to Definite Descriptions (mean = 37.0%; Tukey's HSD, p = 0.048) but not compared to Proper Names (mean = 42.3%; Tukey's HSD, p = 0.11). Also due to the non-significant difference between Proper Names and Definite Descriptions, Antecedent Subjecthood fails to account for our Referential Type results. Third, Referential Competitor – the proportion at which there was another explicitly mentioned character in the previous scene description matching in gender and number with the referring expression in question – was significantly higher for Long Intervals than for Short Intervals (mean = 50.1% vs. 39.9%, respectively; *F*(1,388) = 4.48, *p* = .035). Moreover, Referential Competitor was also significantly higher for Proper Names (mean = 68.0%) compared to Pronouns (mean = 15.7%; Tukey's HSD, *p* < .001) and Definite Descriptions (mean = 51.4%; Tukey's HSD, *p* < .001), and significantly higher for Definite Descriptions compared to Pronouns (Tukey's HSD, *p* < .001), thus posing the most serious contender to Interval as the critical factor in our Referential Type results. However, Referential Competitor was uniform across Intervals for Pronouns (16% for LI and 15% for SI), thus providing no explanation for the divergence between singular- and plural-denoting pronouns (see Figure 6).

The Interval effect on Referential Type frequency was most prominent in the Proper Name category, which consisted mainly of references to the main characters (the boy, the frog, and the other pets). Most of the secondary or minor characters (e.g., the lady eating salad, the valet, and the band members) were not given a proper name, and were not mentioned as frequently as the main characters. Although character prominence was thus an important factor in proper name assignment (Sanford et al., 1988), it cannot explain the Interval effect within a Referential Type, such as Proper Names or Singular/Plural Pronouns.

3.3. Discussion

Narrators preserve content structure in their linguistic encoding of story time, or intervals between events. In addition to 'referential distance' (Givón, 1992) and other text-based factors of accessibility, the 'psychological' distance in mental representation due to a gap in story time in our study is another scale of relevance to referential accessibility in narrative production. A long interval between events reduces the accessibility of referents and increases the need for a strong, specific cue to bring referents back in focus. Our results demonstrate that narrators use proper names rather than anaphoric devices to fulfill this role. In contrast, shorter intervals without a shift in situational representation are more felicitous for anaphora with definite descriptions and singular pronouns. In short, not all scene transitions or temporal gaps are treated in the same way in narrative production; rather, magnitudes in story time are represented to have observable impact on the narrator’s discourse organization.

3.3.1. Situation Models in Discourse Production

Our temporal marking results (in Chapter 2) and referring expression results (in this chapter) demonstrate the impact of nonlinguistically conveyed intervals on narrative production, and thus reveal the pervasive impact of situation models in discourse. Linguistic encoding of content structure was present even though the participants were told that their addressee, an imaginary child, would see the picture books as their stories are read to her/him; that is, narrators provided linguistic cues that would help the addressee bridge long intervals even though the intended referent was obvious from nonlinguistic context. Model-based accounts of discourse comprehension (van Dijk & Kintsch, 1983; Johnson-Laird, 1983; Garnham, 1991; Zwaan, 1999) can thus be extended to discourse production, and interlocutors cooperate to align their situation models (Pickering & Garrod, 2006).

3.3.2. Temporal Shift vs. Other Factors

Although we don’t deny the importance of other salience factors that have been found to influence referential choice, Interval seems to be the best explanation for our data. Alternative factors of Textual Distance, Antecedent Subjecthood, and Referential Competitor find partial support in our data, but fail to account for our Interval effects on Referential Type frequencies.

Vonk et al. (1992) discussed their referential type results in relation to *thematic* or topic continuity/discontinuity as a global dimension of coherence, in which “[a] thematic shift can consist in a change of time or place” (p. 331). Although topic continuity is conceptually related to temporal continuity, there are reasons to think that our results don’t just reflect topic continuity/discontinuity. First, time is distinguishable from other situational dimensions such as space and protagonist (Magliano, Zwaan, & Graesser, 1999). To test whether time is distinguishable from theme as well, we obtained scene-by-scene topic continuity estimates from twelve new judges (see Section 2.3.4), and the topic continuity judgments did not show a significant correlation with time interval estimates. Furthermore, regression analyses pitting topic continuity against temporal continuity as predictors of Referential Types revealed that time interval estimate ranks accounted for more variance overall in the frequencies of referring expressions than topic change estimate ranks did (for Proper Names, $R^2 = .15$ vs. $R^2 = .005$, respectively; for Definite Descriptions, very low predictive power for both; for Pronouns, $R^2 = .23$ vs. $R^2 = .22$, respectively).

3.3.3. Discourse Roles of Referential Types

Noting that pronouns were more common in their data after sentence-initial segmentation markers (e.g., *The following day, After the lessons were finished,* and *At the border*) than when such markers were not present, Vonk et al. (1992) argued that there is a division of labor between over-specific referring expressions and sentence-initial segmentation markers in cuing discourse shifts. In our data, however, Referential Type patterns generally held regardless of the presence of lexical or phrasal segmentation marking. This suggests that interval duration has a pervasive effect on both explicit temporal marking *and* referential choice.

With regard to the relative roles of the different Referential Types, our results suggest that proper names re-mentioning a character typically indicate temporal discontinuity; singular-denoting pronouns indicate temporal continuity; and definite descriptions lie between these two categories, consistent with earlier proposals for an accessibility hierarchy (Ariel, 1990; Givón, 1992; Gundel et al., 1993). Between temporal continuity and discontinuity, definite descriptions lean toward continuity marking, as expected for an anaphoric device whose form typically presupposes an antecedent.

Contrary to some prior literature in which role descriptions were often grouped together with proper names into ‘full NPs’ as segmentation markers (Sanford et al., 1988; Vonk et al., 1992), our data suggest that proper names have a distinct discourse role from definite descriptions, not only for introducing a character of primary importance but also for strongly signaling a discontinuity in the timeline of the content. It is thus important to note the distinction between character introductions (first mentions in a story) and re-mentions, and my results and conclusions regard the latter. Character introductions, on the other hand, are invariably discourse-new (aside from visual cues), and sensitive to character prominence in the story (Sanford et al., 1988) rather than the degree of referential accessibility as a function of situational continuity.

3.3.4. Singular vs. Plural Pronouns

Pronouns are usually considered to serve a 'reference maintenance function' (Sanford et al., 1988). Our results clearly show, however, that not all pronouns serve the same discourse function. Singular-denoting pronouns, whose referents are by nature clearly individuated, behaved as predicted by the accessibility hierarchy (e.g., Ariel, 1990), appearing mostly in temporally continuous events. Most of the pronouns that followed a temporal discontinuity, on the other hand, were plural-denoting pronouns (including the grammatically singular quantifiers *everyone/everybody*) that made a generalization about a group of (three or more) loosely-individuated referents from a previous scene (e.g., ***They*** [the family] *all sat at the table* and *Then* **everyone** [the boy and his many pets] *went on a trip*). This use of plural-denoting pronouns – close in meaning to 'whomever I mentioned before' – usually allows for pragmatic laxity. All of the 44 occurrences of plural-denoting pronouns after Long Intervals except one (*Timmy's father sent* **them both** *to their room*) were such 'generalizing' cases, whereas six of the seven occurrences after Short Intervals referred to just two characters (the boy and the dog) with clear individuation and no room for an exception. The pragmatic laxity and lack of strict individuation associated with the generalizing use of plural-denoting pronouns lead us to believe that these subclasses of pronoun use should be distinguished in an account of their anaphoricity, or sensitivity to accessibility of referents. Specifically, generalizing cases of plural re-mentions involve only a weak or loose tie to their antecedent, and are associated more closely with narrative discontinuity than with continuity. In the future, we plan to follow up on this issue by investigating referent identification and certainty judgments for singular- vs. plural-denoting references in narratives.

3.3.5. Implications

Our finding lends support to McCoy and Strube's (1999) similar intuition in computational linguistics. In developing their referring-expression generation system, they used topic time change in discourse as a major predictor of referential type. Although they limited the application of this time-based rule to coreferential expressions *across* sentences over a *short* text span (to avoid intra-sentential anaphora and limit referential distance), these restrictions match our selection criteria for critical referring expressions closely (see Section 3.2). Gaining further insight into the impact of time change in content on referential choice can thus lead to a predictive model of referring expressions for applications in computational linguistics as well.

Existing "Frogbook narrative" corpora from studies in Berman and Slobin (1994), as well as Chafe's (1980) Pear Stories, can be revisited for further analysis of the relationship between situational dimensions and referring expressions, especially across age groups for a developmental study or across languages for a crosslinguistic study.

Our study does not address an important question: Did the narrators in our study deliberately provide linguistic cues to help their reader detect discourse segment boundaries, or was their choice of referential type an unconscious reflection of differences in their own attentional state? In future studies, we plan to address this issue of 'signal vs. trace' (Bestgen, 1998; Dell & Brown, 1991) with direct manipulation of shared visual access between a narrator and an addressee.

Chapter 4

The Frogbook Project: A Computational Approach to Analysis of Referring Expressions in Narrative Discourse

In the previous chapter, I used manual coding to investigate how the duration of an inter-event interval in story time affects the type of referring expression a narrator uses. In this chapter, I discuss successes and failures of techniques in computational linguistics for a study of how intervals in story time affect the narrator's use of different types of referring expressions. The success story shows that a conditional frequency distribution (CFD) analysis of proper nouns and pronouns yields results that are consistent with our results in the previous chapter, namely, that the narrator's choice of referring expression is sensitive to the amount of time that elapsed between events in a story. Unfortunately, the less successful story indicates that state-of-the-art automatic coreference resolution systems fail to achieve high accuracy for this genre of discourse. Fine-grained analyses of these failures provide insight into the limitations of current coreference resolution systems, and ways of improving them.

4.1. Introduction

4.1.1. Motivation

The theoretical issues discussed in Chapter 3 were addressed with naturally occurring discourse, or corpora. Even with the relatively small size of our collection of narratives, manual coding is extremely expensive in terms of time and human labor. Addressing the issue of the effect of inter-event intervals on referential choice on a large scale requires accurate automatic methods for identification of Referential Types and coreference resolution for the narratives. In this chapter, we first present a simple computational method for analyzing the entire scene descriptions after the Long and Short Intervals to study how inter-event intervals affect referential choice, focusing on Proper Nouns and Pronouns. Our results from the automatic methods are consistent with the results obtained using manual coding of the critical sentences. Second, we present an annotation study of nine narratives with coreference chains. Third, we discuss the performance of two state-of-the-art coreference resolution systems on a sample of our data.

4.1.2. A Brief History of Automatic Coreference Resolution

In computational linguistics, the increasing availability of annotated coreference corpora has led to developments in machine learning approaches to automatic coreference resolution. The task of automatic NP coreference resolution is to determine "which NPs in a text […] refer to the same real-world entity" (Ng, 2010, p. 1396). Successful coreference resolution often requires real-world knowledge of public figures, entity relationships, and aliases, beyond linguistic parameters such as number and gender features.

Early models of coreference resolution were based on local binary classification deciding whether a mention and an antecedent candidate co-refer or not, independently for each pair without taking into account any previous coreference decisions or context (Soon, Ng, & Lim, 2001; Ng & Cardie, 2002). Because each mention pair or pair instance in this approach is assumed to be an independent event, mention-pair models fail to capture the competition among mentions for the best antecedent candidate. Yang, Zhou, Su, and Tan (2003) proposed 'twin-candidate' classification as a way of relating mention pairs, to pit two candidate pairs against each other in a tournament fashion to choose the best candidate. Even this approach, however, is not fully global as there is no simultaneous consideration of all the mentions in previous text.

The independence assumption of the mention-pair models also leads to their inability to capture transitivity in coreference, and thus the need for a separate clustering algorithm for combining pairwise coreference decisions. Further, mention-pair models must specify how training instances are selected, because using all mention pairs leads to a skewed distribution with a much greater number of negative (non-coreferential) instances and is thus undesirable (Ng, 2010).

In order to better represent the global nature of coreference resolution, scholars have proposed entity-mention models (e.g., Luo, Ittycheriah, Jing, Kambhatla, & Roukos, 2004; Yang, Su, Zhou, & Tan, 2004), in which coreference *clusters* or *chains* are considered instead of antecedent *mentions* as antecedent candidates. These models employ cluster-level features, such as ANY and ALL. Unfortunately, despite the advantage in expressiveness that clusters provide, entity-mention models have not outperformed mention-pair models.

More recently, Rahman and Ng (2009) developed a cluster-ranking model, which considers all pairs of a mention plus a preceding cluster simultaneously when assigning values to the preceding clusters in competition.

Coreference resolution can be improved by using compatibility in number and gender, selectional preferences of verbs, and relations among nouns (based on WordNet or Wikipedia). Ranking approaches (e.g., Rahman & Ng, 2009; Denis & Baldridge, 2008) have generally outperformed their binary-classification counterparts (e.g., Soon et al., 2001; Ng & Cardie, 2002). Finally, Ng (2010) and Lee et al. (2011) argued that joint learning of mention/discourse-newness detection and coreference resolution improves model performance, whereas Denis and Baldridge (2008) and Haghighi and Klein (2009) advocated modularity of representation of linguistic features so each can be improved independently.

4.2. Interval Effect on Referring Expressions: A Basic Computational Approach

In order to address the question of how inter-event intervals in a story affect the narrator's choice of referring expressions, we used the Natural Language Toolkit (NLTK; Bird, Loper, & Klein, 2011) to develop a scene-based text segmentation tool specialized for narratives like ours (see Bird, Klein, & Loper, 2009, Section 6.2, for the sentence segmenter 'segment_sentences'). For our purposes in analyzing the impact of perception of scene transitions, scene boundaries were more important than sentence boundaries.[14]

We analyzed the frequency of Pronouns and Proper Nouns that refer to characters in the entire scene descriptions following the Long and Short Intervals with CFD (conditional frequency distribution) tabulation.[15] We expected *entire scene*-based frequencies of Referential Types to replicate the Interval Effect in Chapter 3 based on manual coding of just critical sentences (the first sentence after a Long or Short Interval) and first re-mentions. The results in Table 3 are consistent with our previous results (see Table 2 in Chapter 3). The 'Long Interval' (LI) scenes and the 'Short Interval' (SI) scenes

[14] Unfortunately, NLTK's POS tagger used in the pre-processing is not perfectly accurate. For example, it mislabels a commonly used name *Billy* for the boy character as an adverb (probably due to the *-ly* ending). It would be better to import the parsing output from the preprocessing of advanced coreference resolution systems, for systems that generate either separate output files for different modules or output files that are easy to manipulate so that only the relevant results can be extracted.

[15] We did not look at Definite Descriptions, because they include inanimate objects and other non-characters, and we have not developed an automatic tool for filtering these out from analysis.

show opposite patterns in relative frequencies of our target part-of-speech tags – Pronouns (nominal (PRP) and possessive (PRP$) forms) vs. Proper Names (NNP).

Table 3

Scene-based Frequencies and Relative Proportions of Pronouns vs. Proper Names after Long and Short Intervals

Book	Scene #	PRP	PRP$	NNP	Total
Frog Goes to Dinner	4 (LI)	62 (27.68%)	56 (25.00%)	106 (47.32%)	224
	5 (LI)	54 (28.88%)	37 (19.79%)	96 (51.34%)	187
	21 (LI)	87 (32.58%)	60 (22.47%)	120 (44.94%)	267
	9 (SI)	45 (47.87%)	22 (23.40%)	27 (28.72%)	94
	13 (SI)	50 (34.72%)	44 (30.56%)	50 (34.72%)	144
	14 (SI)	40 (39.60%)	21 (20.79%)	40 (39.60%)	101
One Frog Too Many	8 (LI)	33 (27.27%)	33 (27.27%)	55 (45.45%)	121
	19 (LI)	63 (32.31%)	42 (21.54%)	90 (46.15%)	195
	20 (LI)	60 (33.90%)	29 (16.38%)	88 (49.72%)	177
	3 (SI)	70 (23.89%)	65 (22.18%)	158 (53.92%)	293
	15 (SI)	69 (35.94%)	50 (26.04%)	73 (38.02%)	192
	23 (SI)	1 (20.00%)	2 (40.00%)	2 (40.00%)	5
Frog, Where Are You?	2 (LI)	89 (29.47%)	70 (23.18%)	143 (47.35%)	302

	3 (LI)	70 (23.89%)	65 (22.18%)	158 (53.92%)	293
	18 (SI)	64 (31.07%)	56 (27.18%)	86 (41.75%)	206
	19 (SI)	63 (32.31%)	42 (21.54%)	90 (46.15%)	195

One can observe that there are generally higher frequencies of Proper Names for the scenes after the Long Intervals compared to the Short Intervals, not only in absolute number but in relative proportion to Pronouns as well. As was the case for hand-coded results based on critical sentences and first re-mentions only, automatic coding of entire scene descriptions indicated that Proper Names were more commonly used after Long Intervals than after Short Intervals, and Pronouns were more commonly used after Short Intervals than after Long Intervals ($\chi^2(1) = 9.50$, $p = .0021$).

A noticeable exception, Scene 3 of *One Frog Too Many* (Mayer & Mayer, 1975), is a very early scene in the picture book, with many character introductions and much discourse-newness (Prince, 1981). (In this scene, a second frog appears for the first time in the picture book, and the first frog, being jealous, is presented in isolation from the rest of the pets for the first time.) This exception suggests that excluding the first few mentions in a coreference chain from automatic analysis may reveal a stronger effect of Interval on the referential type of a re-mention (although one mention for introducing a character does not always establish discourse-givenness from the narrator’s perspective; see Clancy, 1980). Successful automatic coreference resolution would facilitate this kind of analysis.

4.3. Annotation of Referring Expressions in Narratives of Picture Books

To test the performance of coreference systems automatically with gold standards and train coreference resolution systems on our genre in the future, we annotated nine narratives manually with coreference clusters or chains (three narratives for each of the three pictures books, with each narrative written by a different writer). Only animate entities, or characters in the stories, were considered. We used the MMAX2 annotation tool (Müller & Strube, 2006). A coreference schema is available from the Heidelberg Text Corpus (HTC, Malaka & Zipf, 2000) sample directory included in the MMAX2 package. The HTC schema allows marking a mention in terms of the discourse entity or coreference chain it corresponds to, as well as 'np_form' (what type of (pro)nominal it is), 'grammatical_role' (subject/object/other), and 'semantic_class' (abstract/human/physical object/other). We imported the HTC schema to annotate the mention level in terms of coreference, and also created a 'scene' level for our picture-book narratives.

The narratives were independently annotated by the author and a second judge. Because the referents were very clear in the narratives for the picture books, there was only one case of initial disagreement in the authors' coreference decisions, and this disagreement was resolved by discussion. Table 4 shows statistics related to these nine narratives.

Table 4

Descriptive Statistics for Nine Hand-annotated Narratives

Story	Narra-tive	# of Mentions	# of Chains	# of Words	Longest Chain	Average Chain Length	Den-sity
One Frog	1	65	8	280	22	8.13	.23

Too Many	2	71	5	277	29	14.20	.26
	3	52	7	268	15	7.43	.19
Frog, Where Are You?	4	128	13	562	60	9.85	.23
	5	62	12	256	20	5.17	.24
	6	78	11	383	25	7.09	.20
Frog Goes to Dinner	7	271	23	1109	58	11.78	.24
	8	111	21	514	38	5.29	.22
	9	167	26	834	37	6.42	.20

As shown in Table 4, the densities of referring expressions in the nine narratives were uniformly very high (mean = 22% of word tokens in a narrative). The three books differed dramatically in terms of the number of coreference chains, with *Frog Goes to Dinner* having on average almost twice as many coreference chains as *Frog, Where Are You?*, and three times as many coreference chains as *One Frog Too Many*. This difference reflects differences in the content of the three picture books, as evidenced by the fact that, for each of the three books, the three writers had similar numbers of coreference chains.

4.4. Performance of Coreference Resolution Systems on Narratives

We chose two state-of-the-art coreference resolution systems to test: Stanford's Multi-Pass Sieve Coreference Resolution System (Lee et al., 2011) (henceforth, Stanford dcoref) and ARKref (O'Connor & Heilman, 2011). Stanford dcoref consists of an initial mention-detection module, the main coreference resolution module, and post-processing. In this system, global information about the text is shared across mentions in the same cluster in the form of attributes such as gender and number. This system received the highest scores at a recent CoNLL shared task (Pradhan et al., 2011), which Lee et al. (2011) attributed to the initial high-recall component (in mention detection) followed by high-precision classifiers in the coreference resolution sieves. ARKref, the second system, is a syntactically rich, rule-based within-document coreference system very similar to Haghighi and Klein's (2009) system.

We analyzed the performance of these systems on a sample narrative (for *Frog Goes to Dinner*). We expected automatic coreference resolution systems to show poorer performance when applied to our written narratives than that reported in the literature, because most of these systems have been trained on newswire, blog, or conversation corpora, which, although quite heterogeneous, are not similar to our written narratives. Some of the most noteworthy particularities of our written narrative collection include (a) fictional content, in which animals occur frequently and are greatly anthropomorphized, (b) an imaginary target audience of a limited age range (six- to eight-year-olds), and (c) clear scene-by-scene demarcation in the writing process, with a new text input box for each new scene in a picture book. The first point, in particular, may limit the utility of named entity recognition (NER) and WordNet relations among nominals in the preprocessing steps prior to coreference resolution. As we discuss below, preprocessing errors in parsing and NER did in fact contribute to coreference precision errors.

Our written narratives had a lot of singleton mentions for secondary characters and plural combinations of characters. We thus evaluated the performance based on the B^3 measure proposed by Bagga and Baldwin (1998), rather than the link-based MUC (Vilain, Burger, Aberdeen, Connolly, & Hirschman, 1995). We computed the B^3 with equal weighting for all mentions. Stanford dcoref achieved B^3 scores of 0.78 in precision, 0.43 in recall and 0.55 in F_1, while ARKref scores were 0.67 in precision, 0.45 in recall, and 0.54 in F_1. Stanford dcoref includes a post-processing module in which singletons are removed, which partially contributes to the low recall score for the system.

4.4.1. Qualitative Analysis of Coreference Output

In this section, we discuss the errors from both ARKref and Stanford dcoref in depth. The coreference outputs from both ARKref and Stanford dcoref demonstrate that preprocessing errors can lead to errors downstream in coreference resolution. Misparsing is one of the serious issues. For example, in ARKref's output for our sample narrative, the third-person singular verb *waves* in *Billy waves goodbye* (Scene 6) and *Froggy waves goodbye* (Scene 7) was misparsed as a plural nominal and thus a headword of a mention for a discourse entity, and these two instances were marked as coreferent.

A few surprising errors in the ARKref output include (a) marking *the woman* and *him* in the same clause as coreferent despite the gender mismatch, and (b) leaving *the lady* as a singleton and starting a new coreference chain for *her* in the same clause. It is strange that the explicitly anaphoric pronoun did not lead ARKref to link it to the successfully identified mention *the lady*.

Other noteworthy errors common to both systems' outputs were the following:

(1) inconsistent mention detection and coreference resolution for mentions of the frog character with *Froggy*;

(2) failure to recognize cataphora in *Without knowing Froggy's in [his]$_i$ saxophone, [the saxophone player]$_i$ tries to blow harder*… and linking the pronoun *his* to *Froggy* instead;

(3) starting a new coreference chain at Scene 4 at the mention of *Billy*

when the referent (the boy) has been already introduced as *Billy Smith* in Scene 1;

(4) the same type of error for another character (the frog) at an indefinite NP *a frog* in *She is so shocked that there is a frog in her salad*.

With regard to error (1), preprocessing results in the Stanford dcoref output reveal some NER errors in which *Froggy* was mislabeled as an 'organization,' which, along with the absence of *Froggy* in the name gazetteer for the system (Lee et al., 2011), would lead to both precision and recall errors for *Froggy*, as we observed. Error (2) also remains a challenge to current systems.

Error (3) reveals the potential pitfall of overreliance on headwords for mention/discourse-new detection, which leads these systems to miss the internal structure to people's names – namely, [first name + last name] for the same person,[16] which then can be re-mentioned using just the first name. Although in news articles and other formal writing it is typical to mention a person by the last name (e.g., *Obama* rather than *Barack*) as long as the referent is clear, stories, conversations, and other less formal genres would make more frequent use of first names of individuals for re-mentions compared to other genres. Because the importance of coreference resolution is not limited to formal writing, coreference resolution systems need to incorporate name-specific knowledge, either in preprocessing stages such as parsing and NER or in coreference resolution after the preprocessing.

Error (4) is not as undesirable as the other ones: Even for a human annotator, it is more difficult to make a coreference decision for a case like this one, in which the fact

[16] Application to East Asian languages would need to adjust to the opposite 'family name + given name' sequence, often even in English transliteration (e.g., *Kim Jong-il*).

that the salad-eating lady was shocked would come about similarly for any frog, not just Froggy. Although there does not seem to be a rule in these systems for automatically classifying an indefinite NP as denoting a new entity, training on a large corpus would lead to such a tendency because indefinites usually do indicate discourse-newness introducing a new discourse referent.

There were also some surprising successes. In another author's narrative for the same picture book, there were two definite NPs (*the woman* and *the waiter*) for which the definiteness was due to the visual availability of the referent in the scene or a bridging inference (restaurant – waiter; Haviland & Clark, 1974) rather than a previous mention. Definiteness may lead coreference systems to prefer assigning the mention in question to an existing coreference chain rather than creating a new chain, but ARKref processed both of these possibly misleading definite NPs successfully by creating a new coreference chain – perhaps in default of a matching coreference chain – and Stanford dcoref got one right and made a recall error for the other. On the other hand, referring to different minor male characters similarly as *the man* did lead to a spurious coreference chain linking all of these mentions.

4.5. Conclusion and Future Directions

With the NLP tools discussed above, possibilities abound for interesting research on narratives. Based on scene-based segmentation of narratives written for fixed target picture sequences, one can collect various kinds of linguistic and nonlinguistic data associated with the picture sequences and conduct regression analysis to see which factors have the most predictive value for linguistic variation such as Referential Type choice (as well as explicit temporal marking). Important factors to test include temporal and thematic (dis)continuity in the target content (McCoy & Strube, 1999; Vonk, Hustinx, & Simons, 1992) and discourse salience factors (Prince, 1981).

Although the current state-of-the-art systems for automatic coreference resolution still need a lot of improvement, particularly for a narrative genre like ours, the partial success they do achieve may facilitate manual coding. Manual coding can be performed in a human post-processing correction procedure on the output from automatic coreference resolution (M. Flor, personal communication, June 4, 2012). To improve the performance of current automatic systems, we plan to use 'semantic_class' attributes and features such as ANIMACY in the HTC schema as our task-specific filters for selecting just story characters. Moreover, we plan to explore other state-of-the-art coreference systems such as CherryPicker (Rahman & Ng, 2009) and BART (Versley et al., 2008). The NLP tools and techniques discussed above can be applied to cross-document coreference resolution as well (see Bagga & Baldwin, 1998, for discussion of a meta document), although training the systems for narratives like ours would involve much more manual annotation and supervision, particularly because different authors usually assign different names to a given character. In order to limit the amount of manual annotation for training, unsupervised methods for coreference resolution (Ng, 2008; Poon & Domingos, 2008; Haghighi & Klein, 2007) could be used; however, this would require a much larger number of human-produced narratives.

Coreference is far from a simple phenomenon, both for theory and application. Nevertheless, ultimately it would be desirable to improve the automatic coreference resolution systems in ways that reflect corpus-linguistic and psycholinguistic findings, such as referential distance effects (Givón, 1992), and the privileged status in memory of discourse entities in the immediately preceding clause (Clark & Sengul, 1979). The eventual goal is to represent as many of the interacting factors in referential choice as possible, with a weighting scheme or a ranking algorithm sensitive to these multiple factors.

Chapter 5

The Two-Minute Mystery Project: Granularity of Temporal Representation

In this chapter, I continue to explore the impact of temporal (dis)continuity in situational content on discourse representation, but focus on the addressee's side of the discourse. I will review previous literature reporting a narrative time shift effect (Zwaan, 1996) in discourse comprehension. The behavioral evidence reviewed here for temporal magnitude representation in discourse processing is particularly interesting because the relevant experimental manipulation is not of physical time in the real world, as with inter-stimulus intervals, but of linguistic descriptions about fictitious time in an imagined world. A purely text-based account of discourse comprehension that denies situational magnitude representation would not predict this phenomenon.

I will also report the results of my Two-Minute Mystery experiment, a self-paced reading study with probe recognition tasks in which I used multiple levels of duration adverbials in the same unit (*10 minutes/20 minutes/40 minutes later*) and measured the effect of duration on reading times and probe recognition latencies.

5.1. Introduction

As reviewed in Chapter 3, Anderson, Garrod, and Sanford (1983) found that reading a temporal adverbial of long duration reduced the reader's probability of re-mention in passage continuation of a secondary character, especially with a pronoun. Following Anderson et al.'s early finding of a time-based episodic shift in narrative processing, recent studies reported a 'narrative time shift effect' in processing speed and accessibility of prior information (e.g., Zwaan, 1996; Rinck & Bower, 2000; Speer & Zacks, 2005). Zwaan (1996) tested experimentally whether there is default incremental progression in narrative time in discourse representation and extra computation in case of alternative time marking (Dowty, 1986). Zwaan (1996) found that narrative time shifts with adverbials such as *An hour later* and *A day later*, whose intervals exceed the interval of default incremental progression (with expressions such as *A moment later*), led to increased reading times and longer response latencies in recognizing probe words from text prior to the temporal adverbials. Zwaan concluded that any time movement in discourse beyond a contiguous subsequent time constitutes a narrative time shift and requires additional cognitive effort, regardless of the typical duration of the target episode (cf. Anderson et al., 1983, for a scenario-based account of temporally triggered *episodic* shifts in narratives).

Zwaan (1996) and Kelter, Kaup, and Claus (2004) dubbed these findings the 'temporal distance effect,' and Rinck and Bower (2000) also made an analogy to a similar finding in the domain of space, namely, the spatial gradient of accessibility or spatial distance effect. In most studies reporting a narrative time shift effect, however, there is only an all-or-none effect – that is, a coarse-grained two-way division between 'contiguity' vs. 'non-contiguity' (an important exception is the event-related potential (ERP) study by Ditman, Holcomb, & Kuperberg, 2008, discussed below). These studies thus do not tell us much about the extent of the temporal distance representation, or the granularity of magnitude representation.

5.1.1. Discrete vs. Continuous Scales: Experimental Evidence

Evidence for gradient magnitude representation, or a *symbolic distance effect,* has been found in various task domains of cognition (see Chapter 1). In discourse processing, longer distances in story time as indicated by adverbial phrases have been shown to lead to longer reading times and longer latencies in probe recognition. In addition to studies discussed earlier in the chapter, Rinck and Bower (2000) found that when participants learned a building map indicating what object was in what room and read passages about a character moving through the rooms, they took longer to recognize matching object-room pairs after temporal intervals of 'hours' in the story (e.g., *After two hours*) than after those of 'minutes' (e.g., *After ten minutes*). The task involved learning the map in a 30-minute-long self-training session before reading passages about the map, and it thus explicitly triggered a situation model with spatial distances. There are several noteworthy aspects to Rinck and Bower's (2000) results. First, participants did not take longer to respond to a probe when there were more intervening sentences between the target word and the probe (and there was thus more *physical* time intervening between the two), whereas longer intervening *story* time had a robust effect of slowing down response latencies. Second, the object-room pairs used as test probes were locations along the character's path of motion from the initial source room to the final goal room that were not explicitly mentioned earlier in the passage but only implied by the path and the spatial configuration. Thus, the probe results had no textual basis at all, and only a representation of the nonverbally presented configuration can account for them. Finally, the "shorter" distance condition in this study involved 'minutes' (ranging from *a minute* to *ten minutes*), unlike in Zwaan's (1996) study using *a moment*.

Rinck and Bower's finding of a latency difference between 'minutes' and 'hours' conditions raises questions about whether there is a discrete boundary dividing narrative time continuity and discontinuity, and if so where exactly the boundary lies. It is unclear why a continuity-discontinuity boundary should lie somewhere between ten minutes and an hour; moreover, Zwaan's (1996) results showed a general increase (though non-significant) in reading times and response latencies from his 'intermediate' to his 'distant' condition, not just from his 'close' and to his 'intermediate' condition. These considerations encouraged us to investigate the temporal distance effect in a more fine-grained way, with multiple levels beyond a minimal temporal advance in a narrative (see the next section).

In a later study, Speer and Zacks (2005) extended the earlier accessibility results in probe recognition latencies to 'availability' results in probe recognition accuracy using *a moment later* vs. *an hour later*, suggesting that prior information is not just less accessible after a longer interval in story time but often completely forgotten even when participants are given enough retrieval time. Furthermore, Speer and Zacks (2005) demonstrated that the long interval, *an hour later*, itself was sufficient to lead to longer reading times compared to the short interval, *a moment later*, regardless of antecedent retrieval effort for coreference resolution.

Kelter et al. (2004) enriched the discussion with results from their German study. They used action descriptions of different typical duration ('baking cookies' for Long duration vs. 'putting cookies on plates' for Short duration) without temporal adverbials, and demonstrated that reading times for sentences with an anaphoric reference and probe recognition latencies were longer after Long-duration actions than after Short-duration ones. In a separate self-paced reading experiment using duration manipulation with adverbials instead of action types (a two-by-two design, e.g., *for an hour/six hours* vs.

after an hour/six hours[17]), Kelter et al. (2004) found a Duration effect on reading times only for sentences with a durative adverbial (*for*) describing the duration of an intervening event, and not for those with a 'time-shift' adverbial (*after*). They echoed Zwaan (1996) in attributing the lack of a reading-time difference after their time-shift adverbials to the fact that their intervals went beyond contiguous succession, resulting in a narrative time shift and a new representation for the subsequent discourse (a 'fresh start'). In Kelter et al.'s account, it is only in a single dynamically evolving representation that we would observe effects of duration; in other words, duration is analogously represented only in continuous tracking within a representation and not in a fresh start between representations. Kelter et al.'s finding also adds a twist to the discussion of narrative time shifts in that both levels of duration in their study were at 'an hour' and above and showed differential impact on processing times, contra Zwaan's account, in which anything beyond 'a moment' in duration should result in a narrative time shift and processing effort in a statistically indistinguishable manner.

In an ERP study, Ditman et al. (2008) showed that reading the word *year* in an adverbial *After one year* in discourse evokes a larger N400 amplitude (a sign of difficulty in contextual integration; Kutas & Federmeier, 2000) than reading *hour* in *After one hour*, and *hour* in turn evokes a larger N400 than *second*, beyond what can be accounted for in terms of plausibility of the durations in context. In addition, after these temporal adverbials, overspecified anaphors resulted in a larger N400 for *hour* and *second* than for *year*. Ditman et al. (2008) argued that these N400 effects demonstrate conceptual

[17] Kelter et al. (2004) did not state clearly whether the 'time-shift' adverbials were sentence-initial in their original German stimuli.

integration costs and that these point to a neurophysiological basis for rapid discourse-time tracking.

In sum, recent literature indicates that the processing time effects sensitive to temporal intervals in discourse reflect cognitive effort required for either (a) establishing a new unit of discourse representation after a time shift, or (b) dynamic tracking of topic time within a representation while preserving duration information in a fine-grained way.

5.1.2. Motivation and Hypothesis

Recent discussion in the framework of embodied cognition (Zwaan, 2004; Glenberg & Kaschak, 2002) proposes that tracking the topic time in discourse involves a perceptually grounded symbol analogical to real-world magnitudes (Barsalou, 1999). Thus, more empirical investigation is necessary as to how fine-grained our temporal representation is in tracking topic time. While previous studies have helped accumulate valuable insights, several limitations exist: The studies used either different units of measurement for the levels of experimental manipulation (*moments* vs. *hours* vs. *days*, etc.), or had only two levels of temporal manipulation. The use of different words or units is problematic because it introduces possible confounds, such as word frequency. Moreover, we cannot assume that there is a single scale for standardized time units such as *days*, *months*, and *years* during language use, reflecting the actual size relations (e.g., one year = 12 months = 365/366 days). It is possible that in processing such different time units, we shift gears between different scales in representation so that *two months* may have more similar processing impact to *two days* than to *sixty days*.

In addition, there were many artificial aspects to the text stimuli used in previous experiments (e.g., fixed story length, a fixed rhetorical frame for each sentence position in a story, and excessive disruptions in the reading task with as many as 12 probe trials in a narrative), and it is possible that these affected the results (Speer & Zacks, 2005; Zwaan, 1996; Rinck & Bower, 2000; Kelter et al., 2004).

In our study, we used three levels of story time, all within the same temporal unit of measurement. We also used naturally occurring passages, rather than ones created and explicitly controlled for experimental purposes (see the Method section). Converging evidence for magnitude representation with temporal adverbials and the notion of real-world analogues in mental representation led us to expect more fine-grained distinctions between intervals of different durations than just a two-way distinction between contiguity and a time shift. We thus hypothesized that there would be a gradient effect of duration of intervals in story time on reading times and probe recognition latencies, with a significant difference between our three levels of intervals in the same unit.

5.2. Method

5.2.1. Participants

Thirty-six native English-speaking undergraduate and graduate students at Rutgers (mean age = 21 years; age 18-31 years) participated in the experiment for course credit or monetary remuneration. None of them reported a history of a language disorder.

5.2.2. Stimuli

We used minimally edited versions of 15 naturally occurring mysteries and brain teasers (so-called “two-minute mysteries/brain-teasers”; see Appendix for stimulus texts and sources).

Each experimental story began with a title and contained two probe-word recognition trials, one Main and one Filler trial. All the Main probe trials were preceded by a sentence beginning with a temporal adverbial that was one of three levels (*10/20/40 minutes later*), and required a ‘Yes’ response, with a probe word from the sentence before the manipulated sentence. Although Filler trials weren’t of interest to us, most of the Filler trials were also preceded by a temporal adverbial, *10/20/40 minutes later*, in order to prevent participants from automatically associating a temporal adverbial with a ‘Yes’ response.[18] Each of our 15 experimental stories thus had nine versions (three-by-three Interval combinations between Main and Filler trials). The probe words were usually the sentence-final word, and the probe words in the few exceptions came from earlier in the sentence in order to avoid using plural nouns or pronouns as test probes (see Appendix). All the Filler trials required a ‘No’ response, with a probe word that did not appear in any of our stories (see Appendix). None of the critical sentences except one (Story 7, *Inspector Saunders and the Stranger*, Main trial) contained the critical clue to the mystery.

The Main probe trial appeared first in seven stories, and Filler appeared first in the other eight. The presentation order for the experimental stories was randomly selected from four lists, and levels of Interval for Main trials were chosen on a Graeco-Latin square basis (10-20-40-10-40-20-20-40-10-…) so that each participant saw five instances of each Interval level in the Main trials, and each story was presented at each of the three Interval levels to 12 participants. For Filler trials, we used a random number generator (instead of a Graeco-Latin square) to obtain a pre-determined order for Interval level, in order to avoid a systematic pattern between Main and Filler trials in their level of Interval.

[18] One Filler trial was preceded by a sentence with *10/20/40 days later*, and another Filler was preceded by a sentence with *10/20/40 tables*.

Because we did not artificially control for sentence length or story length, there was a wide range of variation in length across stories. The average experimental story had 12.73 sentences including the title and the mystery question (*SD* = 4.77, min = 7, max = 22). The average sentence across all stories was 12.77 words long (*SD* = 7.12); the average Main trial sentence was 12.53 words long (*SD* = 5.72); and the average Filler trial sentence was 14.87 words long (*SD* = 6.58).

5.2.3. Procedure

Participants read stories on a computer screen. Stories were presented in whole sentence format, and participants advanced to the next sentence by pressing the space bar when they were ready (i.e., self-paced reading in a sentence-by-sentence moving-window paradigm). Each sentence was presented on a new line, and the rest of the story area on the screen (besides the current sentence) was in underscores indicating only the lengths of the invisible sentences. Participants read and solved three practice stories before moving on to 15 experimental stories. All text was presented in 22-point black Tahoma fonts on a white background, center-aligned vertically and left-aligned horizontally. The shortest line in our experimental stories was one word long (a title, *Fishing*, and two queries, *Why?*), and the longest one was 33 words long. No sentence went over a line.

Because a pre-test with five participants revealed that the mystery-solving task was too difficult after reading a story just once without knowing what the mystery question will ask, we gave participants the option to read each story a second time before moving onto the next one. The second presentations of mysteries involved no probe recognition task. In order to prevent participants from spending too much time on a story, we presented a '5-minutes!' warning in case they went over five minutes from the beginning of a story.

Due to the inherent challenge in the task of solving puzzles, all participants were attentive throughout their session and said they enjoyed the task. In debriefing, none of the participants reported knowledge of the structure of our experimental design.

5.2.4. Apparatus

We used a Dell Optiplex 980 desktop computer to present stimuli on a 24-inch widescreen LED monitor (ViewSonic VX2433wm, Full HD 1080p display with 1920 x 1080 resolution), and participants sat at a comfortable distance (typically about one meter) from the screen. At the end of each story, participants typed in their answer on a separate notebook computer showing a text box.

5.3. Results

On average, participants chose to read a story a second time for 10.69 out of 15 stories (*SD* = 3.01, min = 2, max = 15), and the 15 stories were read a second time by an average of 25.67 out of 36 participants (*SD* = 6.87, min = 7, max = 34).

5.3.1. Accuracy in Solving Mysteries

Before discussing our main dependent variables of interest – reading times and probe response latencies – we report accuracy of participants' responses to the mystery questions. Three judges including the author scored the participants' responses to the 15 experimental stories. We required a correct response with adequate reasoning, but we were otherwise lenient because the mysteries were difficult to solve. The three judges assigned uniform scores to 491 out of 541 responses (91.1%), and for responses on which the accuracy decision was split, we took the majority decision as the final score.

Participants’ accuracy in solving mysteries was generally low (mean = 41.9%), with the best-performing participants solving 11 of the 15 mysteries successfully, and the worst-performing participant solving none correctly. The primary purpose of the mystery-solving task was to make sure participants paid attention to the content, and as even the incorrect responses indicated participants’ recall of relevant story content, we did not exclude anyone’s data from the reading time and probe latency analyses below. The most difficult story was *Ready for Bed* (see Story 3 in Appendix), which only three of our 36 participants solved successfully, and the easiest story was *A Drink for a Crow* (see Story 13 in Appendix), which 29 participants solved successfully.

5.3.2. Reading Times

We analyzed the sentence reading times for first and second readings separately, because readers in a second reading may either strategically look for clues to the mystery while ignoring content they judge to be unhelpful for the task (thus reducing an Interval effect), or represent the situational content even more vividly after novelty of information is gone. Separate analyses can address both possibilities.

5.3.2.1. First readings.

We performed linear mixed-effects modeling with crossed random effects for Participant and Story (Baayen, Davidson, & Bates, 2008) on the effect of Interval on raw reading times. We analyzed reading times for sentences before Main probe trials (N = 540). Because we predicted a specific direction of Interval effect, namely, an increase in reading times from *10 Minutes* to *20 Minutes*, and from *20 Minutes* to *40 Minutes*, we report one-tailed t-test results for all our linear mixed-effects modeling below. The mean reading times in Main trials for the three levels of Interval were as follows: (a) *10*

Minutes: mean = 3316 ms, *SD* = 2967 ms; (b) *20 Minutes*: mean = 3581 ms, *SD* = 3106 ms; and (c) *40 Minutes*: mean = 3651 ms, *SD* = 3295 ms. The effect of Interval for all Main trials approached significance (t = 1.50, N = 540, p = .067). Two separate ANOVAs for Main trials – with Interval as a fixed factor in both, and Participant (F_1) and Story (F_2) as a random factor in each separately – showed that the main effect of Participant approached significance (F_1(35, 70) = 1.58, *MSE* = 7.38, p = .053), and the main effect of Story was significant (F_2(14, 28) = 71.22, *MSE* = 3.00, p < .001). All other effects were non-significant in these ANOVAs, with *F*s < 1.9.

Planned pairwise comparisons indicated that the reading time difference between *10 Minutes* and *20 Minutes* approached significance (t = 1.38, n = 360, p = .085), that between *20 Minutes* and *40 Minutes* was non-significant (t = 0.33, n = 360, p = .37), and that between *10 Minutes* and *40 Minutes* approached significance (t = 1.56, n = 360, p = .060).

5.3.2.2. Second readings.

We analyzed raw reading times in second readings for sentences with the critical temporal adverbial (ones that preceded a Main probe trial in first readings) (N = 384 trials). In second readings, the mean reading times were not in the predicted direction for the three levels of Interval (*10 Minutes*: mean = 4286 ms, *SD* = 6683 ms; *20 Minutes*: mean = 4002 ms, *SD* = 5657 ms; *40 Minutes*: mean = 4744 ms, *SD* = 7732 ms), and linear mixed-effects modeling with crossed random effects indicated no main effect of Interval on reading times (t = 0.80, N = 384, p = .21).

5.3.3. Probe Recognition

Participants were quite accurate in probe recognition tasks: All participants achieved accuracy of 80% or higher, except two who were at 73% (mean = 13.67 out of 15 trials, *SD* = 1.22, 91.1% accuracy). Accuracy was high for all stories at 86% or above, except for one story (*Murder at the Inn*) at 69%. Mixed-effects logistic regression with crossed random factors for Participant and Story indicated that there was no Interval effect on accuracy ($t = 0.75$, $N = 540$, $p = .23$). Two separate ANOVAs (F_1 and F_2) indicated that the main effect of Story was significant ($F_2(14, 28) = 3.45$, $MSE = 0.062$, $p = .003$), and the main effect of Participant approached significance ($F_1(35, 70) = 1.53$, $MSE = 0.065$, $p = .065$). In the ANOVAs, all other effects, including the main effect of Interval, were non-significant, with all *F*s < 1.2.

For response latencies, we analyzed only Main trials that participants answered correctly. The mean probe latencies were not in the predicted direction (*10 Minutes*: 2389 ms, *20 Minutes*: 2329 ms, *40 Minutes*: 2335 ms), and linear mixed-effects modeling with crossed random effects for Participant and Story revealed that the Interval effect on response latencies was non-significant ($t = 0.18$, $n = 492$, $p = .43$). Two separate ANOVAs (F_1 and F_2) on Main trials revealed a significant main effect of Participant ($F_1(35, 70) = 2.95$, $MSE = 2.59$, $p < .001$), but no other significant effect, with all other *F*s < 1.6. In sum, response latency data only revealed a great deal of individual variation in speed of recognizing a probe word correctly.

5.4. Discussion

Our reading time results were in the predicted direction, increasing with longer duration of story time intervals, from *10 Minutes* to *20 Minutes* to *40 Minutes*. Although the Interval effect on reading time was non-significant in second readings, it was marginally significant in first readings, with marginally significant pairwise differences in reading time between *10 Minutes* and *20 Minutes*, and *10 Minutes* and *40 Minutes*. We conclude that our reading time data provide some further support for temporal magnitude representation in discourse comprehension. Crucially, we observed the graded effects of Interval on reading time when the temporal adverbial contained an Arabic numeral and a uniform unit of measurement, *minutes* – two novel aspects of experimental design. In second readings, however, participants seemed to have read the sentences more strategically and selectively, not paying close attention to all sentences as in first readings. With different tasks such as those based on comprehension queries instead of mysteries, second readings may also involve situational representation of the discourse content.

Interval levels with larger differences among one another – e.g., *10 minutes* vs. *30 minutes* vs. *90 minutes*, or *10 years* vs. *20 years* vs. *40 years* – may strengthen the Interval effect on processing effort. In manipulating the temporal magnitudes, however, one should be aware of complications that may arise from frequently used alternative expressions (*half an hour* for *30 minutes*, or *an hour* for *60 minutes*) and from natural scales including the diurnal cycle (for hours) and a person's development from childhood to adulthood (for years). Alternatively, using intervening eventualities of different duration between a target and a probe in probe recognition or anaphora resolution tasks as in Kelter et al. (2004) may show a stronger effect of Interval than using just temporal adverbials.

Our probe recognition latencies, on the other hand, did not show any impact of Interval. A more careful procedure for selecting the probes – e.g., equating textual distances from the targets and usage frequencies across items – may be required to improve sensitivity. Another limitation of the study was that although some participants reported in post-task debriefing that they were familiar with a few stories, we didn’t take the factor of prior familiarity into account in our analyses. In the future, we plan to use a single source or more similar sources of naturally occurring text stimuli so the probability of prior exposure is equally low for all stories.

Chapter 6

Conclusion

6.1. Story Time and Situation Models

In discourse, information about time in the world or situation is pervasively marked. It is present in explicit temporal references, grammatical markers such as tense and aspect, and the types of referring expressions. Furthermore, temporal reasoning for discourse representation goes beyond explicitly marked times (Lapata & Lascarides, 2006), involving our knowledge of inherent aspectual classes of eventualities and typical duration of familiar routines.

In this dissertation, I discussed evidence for temporal magnitude representation in the production and processing of discourse that partially preserves differences in duration of intervals and eventualities in the narrative content. Narrators' use of temporal markers and different types of referring expressions in production and readers' reading times for sentences containing time-shift adverbials in comprehension suggest that temporal movements in story time of different magnitudes are not represented as equidistant shifts in a dimension of discourse representation, but have differential impacts on both narrators and addressees. My findings are consistent with Pickering and Garrod's (2006) view of communication as alignment of situation models, and add to the body of evidence that story time is a critical dimension of discourse/situation model.

Evidence for content-based or propositional representation of text was found in early studies of text memory (e.g., Sachs, 1967; Wanner, 1968; Anderson, 1974) and discussed in terms of Chomsky's (1965) deep structure, but subsequent literature has pointed to a much broader base of discourse representation including perceptual context and general knowledge (e.g., Bransford & Johnson, 1972; van Dijk & Kintsch, 1983; Fukumura, van Gompel, & Pickering, 2010). In particular, temporal magnitude effects in discourse have been discussed in the framework of situation models of discourse representation. One account in this framework is Zwaan, Langston, and Graesser's (1995) event-indexing model, according to which we continuously keep track of five core dimensions of discourse content: time, space, protagonist, causality, and intentionality. A shift in one of these dimensions may not result in an overall discourse model shift, but shifts in many of these would, leading to extra cognitive effort for building a new representational unit (see also Gernsbacher, 1990, for shifting based on various dimensions of content). In Rinck and Bower's (2000) account, situation models are a mental map based on an associative network with nodes for spatial locations and links for preserving distances. Kelter, Kaup, and Claus (2004) and Speer and Zacks (2005) similarly argued that situation models are not a series of 'static snapshots,' but rather structured representations reflecting the real world – specifically, a series of dynamically evolving representations that preserve magnitudes in situational dimensions of the content.

These recent developments have encouraged theorizing about embodied or 'action-based' meaning (Glenberg & Kaschak, 2002) and motor resonance (Zwaan, Taylor, & de Boer, 2010), in which symbolic representation is not completely amodal and free of association with particular modules, but has an indispensable sensorimotor underpinning for simulating the perceptual experience of 'being in the described content' (e.g., Zwaan, 2004; Barsalou, 2008; Willems & Casasanto, 2011). Although dynamic tracking of topic time in discourse on a mental timeline need not entail particular bodily states or inferences about potential for actions, my data do suggest that mental representation of magnitudes in situations pervades linguistic processes. Acknowledging magnitude representation in symbolic operations does not detract expressive power from symbols, as magnitude representation is quite domain-general. In language, domains such as time, space, actuality/certainty, and gradable predicates involve degrees on a scale even when they are not used in combination with a number word, and are often grammatically encoded in many languages. Dynamic representation preserving magnitudes in linguistic meaning is not surprising in light of dynamic mental representation in other cognitive domains (see Kelter et al., 2004, for a list of references). Representation of time in discourse may involve a shared system for quantity domains, for which the inferior parietal cortex plays a central role (Walsh, 2003; Brannon & Roitman, 2003; see also Gallistel & Gelman, 2000, for mental magnitudes). The mental timeline for language that I am proposing, however, is not a uniform metrical number line. While the mental representation is based on an underlying magnitude representation system, it is at the same time crucially determined by the discourse context, and linguistic and cultural conventions of talking about time (see Section 6.3.2).

Our findings, particularly the reading time results in comprehension, pose a challenge to linguistic accounts that treat temporal representation as derived from event primitives (e.g., Moens & Steedman, 1988). Although inferences about events do clearly play a major role in the ‘contraction’ (Kamp, 1979), ‘pragmatic coarse-graining’ (Bittner, 1999), and other contextual adjustments in temporal interpretation, the temporal distance effect on reading speed (and accessibility of prior information in other studies) suggests more fine-grained tracking on a mental timeline than non-temporal primitives would predict, and this effect cannot be accounted for in terms of plausibility (e.g., Ditman, Holcomb, & Kuperberg, 2008; Zwaan, 1996). In an event-based account, the impact of story time intervals may be interpreted in terms of inferences about what kinds of events must or might have happened during an interval – with slower reading and recall thus reflecting more difficult inferences for coherence or inferences about more numerous events – but such an approach does not provide a better explanation of the temporal distance effect than an account based on temporal primitives, for which there is independent evidence in other cognitive domains.

I conclude that theories of dynamic discourse representation should incorporate or link up with a fine-grained magnitude system for situational dimensions. The Discourse Representation Structure example in Chapter 1 (Example 5') contains the number '3' for *3 days later*, suggesting the possibility of linking up with number-word semantics – e.g., Fox and Hackl's (2006) dense scale for magnitudes in natural language. Our production findings suggest, however, that interval representation for discourse extends beyond duration specified with number words. It is not that theories of dynamic semantics cannot represent, or are incompatible with, fine-grained magnitude representations. Formal semanticists in early literature tended to focus more on discretization in interpretation (e.g., Kamp, 1979), whereas experimental psycholinguists in more recent literature have addressed gradient behavioral effects in online processing of discourse. Semantic formalisms have begun to incorporate more incrementality, with frequent updates of information states within a sentence (Bittner, 2012), and recent psycholinguistic accounts note parallel mechanisms between comprehension and production (e.g., Pickering & Garrod, 2007). It is a desirable goal to integrate these insights across disciplines in a unified framework of discourse dynamics.

6.2. Relevance to Coherence-based Accounts of Discourse

My findings of Interval effects on discourse production and processing are most relevant to the narrative genre with a coherent story timeline. The majority of the transitions between sentences in our narratives, especially in the first sentence after a critical Long or Short Interval, fall into the 'Occasion' relation in the Contiguity class (Kehler, 2002; Hobbs, 1990), in which "the first event sets up the occasion for the second" (Hobbs, 1990, p. 87). When other kinds of relations do occur in a narrative genre like ours, our few instances in (1)-(3) (after critical intervals) suggest that the Parallel and Contrast relations may usually occur in descriptions of simultaneous actions or states by different characters (see Kehler, 2002, for an example of no temporal constraint in a non-narrative example), whereas the Explanation relation for a single agent may cite a current state (or an earlier event) as a reason:

(1) *While he* [the boy] *was sleeping, Froggy decided to escape*. (Parallel)

(2) *The man falls into the drum because he can't see!* (Explanation)

(3) *Blackie is scared because he doesn't know how to swim, but Roger is OK.* (Explanation & Contrast)

As these examples suggest, descriptions of Parallel or Contrast relations between different characters' actions would require explicit character mentions with proper names or specific definite descriptions (especially when prosody cannot serve as a cue); in contrast, Explanations for the same agent would typically contain a pronoun to maintain the referent. Non-Occasion relations each have unique temporal and referential constraints, especially in relating clauses within a sentence, and they may thus show particular kinds of marking – e.g., for Parallel/Contrast, markers of simultaneity such as *meanwhile* and *at the same time* (see Aksu-Koç & von Stutterheim, 1994). However, when non-Occasion relations involve topic time movements over an interval, these relations are expected to show an Interval effect as well. For example, Explanations involving a temporally more distant flashback may lead to slower processing and more explicit re-mentions compared to a closer flashback (after appropriate preceding context) (e.g., *The boy was so sad because* ***Lola*** *had left* [a week before].[19] vs. *The boy was so sad because* ***she*** *had left* [earlier that day].).

We can expect that building and maintaining a situation model would be more difficult for processing narratives with a lot of temporal disruptions or anomalies, e.g., novels in the stream-of-consciousness mode or in the genre of fantastic realism (Graesser, Kassler, Kreuz, & McLain-Allen, 1998). Genre-specific comprehension strategies may thus exist that involve inhibition of situation models. However, to the extent that situation models can reasonably be constructed for the target content (e.g., for mini-narratives embedded in a globally non-narrative passage), I predict model-based impact – greater cognitive effort required for multiple models or model shifts compared to model maintenance and enrichment.

6.3. Outstanding Issues and Future Directions

6.3.1. Audience Design, or Trace vs. Signal

Patterns in frequencies and types of linguistic expressions such as (pro)nominal references and lexical and phrasal temporal markers may be a reflection of the narrator's

[19] These are my own examples, not from our narratives.

own attentional processes, an intentional signaling device to aid the addressee's comprehension, or both. Clancy (1980) noted this distinction between speaker-based tendencies and listener-oriented strategies, and Bestgen (1998) used the terms, 'trace' (of the narrator's own discourse segmentation) vs. 'signal' (for the addressee's discourse segmentation), to make the distinction. Although the addressee may make use of these patterns in discourse whether they were intended or not, it is important to tease apart 'trace' and 'signal' to gain precise insight into the speaker's intentions and strategic communication. In recent literature (e.g., Barr & Gann, 2011), Clark and Murphy's (1982) notion of 'audience design,' which originally referred to the speaker's tailoring of language production to particular listeners' shared knowledge with the speaker, has been extended to 'generic-listener adaptations' (Dell & Brown, 1991). In the Frogbook studies reported in this dissertation, narrators marked shifts in story time even though they were told that the audience would have full access to the pictures as well. The linguistic patterns we observed in the temporal and referring expressions in our narratives were not intended for providing information that the audience would lack. However, the fact that the audience would have visual access to the content does not eliminate the possibility that our narrators intended these linguistic devices as deliberate, reinforcing signals to aid the addressee's comprehension.

Because my current experiments do not manipulate the addressee's access to the target content while holding the narrator's access constant, in future studies I will experimentally manipulate the level of shared access to content to address the issue of intentionality in linguistic marking of discourse structure.[20] For example, in a future study, the participant will tell a story to a confederate addressee. Half of the time, the confederate will have full access to the pictures that the narrator describes, and half of the time the confederate will not. I predict that narrative sensitivity to the temporal structure of the storyline will be evident in both conditions, but at differing degrees. Structuring devices for discourse timeline do not arise only when the narrator knows the addressee doesn't have enough information to recover the content structure; that is, they seem to reflect the narrator's own situation model to an extent. Even when there is shared nonlinguistic access to the situational content, however, narrators may nevertheless adapt their narrative to reinforce the addressee's structuring of discourse representation (see Barr & Gann, 2011, for a review of both egocentric and audience-centered tendencies in language production).

6.3.2. Pragmatics of Scaling

Clearly, the mental timeline must resort to simplification through compression and discretization; otherwise, a single mention of a century-long interval or movies such as *Back to the Future III* would take more than a lifetime to process. Many scholars (e.g., Dowty, 1986; Webber, 1988; Moens & Steedman, 1988) have noted the default incremental progression of narrative time in the absence of other temporal specifications, and have also emphasized that context determines the level of detail in the discourse timeline. We use our background knowledge of the type of eventualities being discussed

[20] There are recent efforts in this direction based on manipulations of addressee presence (Kantola & van Gompel, 2011) and the addressee's access to visual stimuli (Yoon, Koh, & Brown-Schmidt, 2011).

to decide whether a movement in narrative time is continuous (with simple continuation in the same dominant tense) or discontinuous (with additional temporal marking).

In his discussion of contextually and culturally determined boundaries between time intervals for the so-called metrical tenses, Botne (2012) noted a layering of time scales (hours, days, months, years, or life span, with the exact inventory depending on the language), and argued that deciding proximity vs. remoteness in time is relative to the relevant scale in discourse, and a year is not necessarily more remote than a couple of months as on an absolute linear scale in most of these languages. Linguistic units of time (such as minutes, hours, days, weeks, months, years, etc.) may lead to re-scaling of the mental timeline for the different lexical units so that the mental representation for *90 minutes* does not necessarily involve exactly the same magnitude as that for *1.5 hours*. Also, the pragmatics of round numbers in time (*a quarter-hour*, *a half-hour*, *a week*, etc.; see Jansen & Pollmann, 2001) may affect the mental timeline in discourse so that saying *120 minutes* or *121 minutes* instead of just *2 hours* increases the granularity of the representation due to implicatures.

The model-based representation can also zoom in and out on the discourse timeline based on the scenario or episode under discussion: For instance, a narrative about a marathon would involve a coarser-grained timeline, compared to one about waiting for traffic lights to change with a magnified discourse timeline (J. Musolino, personal communication, June 1, 2012; see also Anderson et al., 1983; cf. Zwaan, 1996). As discussed in earlier chapters, psychologists do not agree about the role of pragmatics in scaling the mental timeline. Anderson, Garrod, and Sanford (1983) have argued that people’s schemas of typical duration of episodes determine continuity vs. discontinuity in narrative time (see also Gernsbacher, 1990; Haberlandt, Berian & Sandson, 1980). Zwaan (1996), on the other hand, has argued that the continuity vs. discontinuity dichotomy depends on a context-independent boundary between a short interval (e.g., a moment) and long intervals (e.g., an hour, a day, etc.). However, although Zwaan's reading time and probe latency results showed only non-significant differences between his 'intermediate' and 'long' interval conditions, the differences were mostly in the predicted direction, even closely approaching significance in one experiment (Experiment 2A, reading times at the temporal marker). My data from a task with three levels of interval duration with the same unit (*10 Minutes*, *20 Minutes*, and *40 Minutes*) showed some effect of interval duration on reading times in participants' first reading of stories, suggesting that temporal distances are not a simple dichotomy between 'a moment' vs. longer intervals. More studies that avoid possible confounds of lexical units (*minute* vs. *hour* vs. *day*) and of round numbers in time are necessary for precise insight into the relationship between target magnitudes in content and cognitive effort required to represent them during discourse use.

6.3.3. Times and Worlds: Temporality and Modality

Languages, such as Hindi, have temporal words or tenses that indicate only the temporal magnitude – e.g., 'one day away from today' – and not the direction – past or future. If magnitudes for time in language are preserved on a mental timeline, and for space in a mental map (see, e.g., Zwaan, 1999; Rinck & Bower, 2000), an interesting issue is the extent of magnitude representation for modality. Linguistically, probability/actuality can be expressed with modifiers (*possibly*, *probably*, *certainly*, etc., and the corresponding forms in other parts of speech) and modal verbs (*may*, *might have*, *will*, *would have*, etc. in English, and inflectional endings in other languages; see Bohnemeyer, 2009; Bittner, 2011; Botne, 2012). It is often inherent in the predicate by lexical entailment (e.g., *John managed to run* entails running, but *John seemed to run* doesn't; see Karttunen, 1971, and Nairn, Condoravdi, & Karttunen, 2006, for implicative strength of various predicates).

Modality in semantics is discussed in terms of possible worlds (e.g., Heim & Kratzer, 1998) or situations (Barwise & Perry, 1983). Translated into situation models in discourse planning and comprehension, degrees of certainty or actuality can be represented by modal indices for the entire situation models. For instance, the most certain or actual events, in the present progressive with direct observation, would have a high index or be 'in bold'; past-tense declaratives less so, due to the fragility of memory; future declaratives generally even less so; and counterfactuals would have a low index or be 'in grayscale' (n.b., counterfactuals have their own range of likelihoods, A. S. Gillies, personal communication, June 1, 2012). Alternatively, one could imagine representation of multiple situation models in parallel with an indexed link indicating the remoteness among one another (see Schaeken, Johnson-Laird, & d'Ydewalle, 1996, for temporal reasoning with multiple mental models – with equal probabilities, in this case).

In semantic literature, the intricate ties between the domains of temporality and modality have long been noted. In his analysis of the present progressive in English, Dowty (1979) noted its dual modal-temporal role, coercing an accomplishment predicate into a statement merely indicating the 'possibility' of completion of the event. Similarly, Botne (2012) noted that some grammatical markers in Sanumá have a dual role for the past tense and degrees/sources of knowledge ('witnessed, verified, supposed'), and that different future tenses in Rugciriku and Kesukuma indicate varying degrees of probability of occurrence (e.g., distinguishing between ‘remote past’ / ‘(typically) yesterday’ / ‘today’ (the deictic center) / ‘(typically) tomorrow’ / ‘remote future’). Efstathiadi (2009) described slight differences in epistemic commitment among modal markers in Modern Greek and their interaction with tense interpretation. The presence of such systematic devices across distant language families for magnitude representation in time and certainty suggests our common potential for complex magnitude representation in language, though through different devices across languages (see also Jaszczolt, 2009, p. 40, for diachronic development of modal and temporal markers).

As an alternative to the common assumption of a single mental timeline, Botne (2012; Botne & Kershner, 2008) has proposed that multiple timelines underlie the use of temporal-modal markers in the languages he discussed. In his account, in addition to what he called the P-domain (an extension into the past and future of the utterance time ‘now’ – intuitively, the actual world), we represent dissociated mental worlds called D-domains, which reflect epistemic or subjective distances. In an alternative view of similarity or remoteness among possible worlds, Dowty (1979) has proposed forward-branching worlds, in which identical histories are compressed into a single line but diverging futures branch rightward in a treelike manner. This account captures in a more intuitive manner the reason why future statements do not have a truth-value at present, but having the truth-value T at the topic future time is possible and having the truth-value F at the same topic future time is also possible, while reducing multiple world-parts with a common state of affairs into one representation.

Although most of the modal markers Botne discussed play a dual role of indicating modality and a temporal direction relative to ‘now’ (past or future), Botne also noted a ‘correlation between temporal distance and modal distance’ in the use of a single remoteness marker for both past and future directions in D-domains in a few languages. Jaszczolt (2009) presented a more extreme view, arguing that temporality *is* (epistemic) modality, with degrees of epistemic commitment or detachment from certainty: The past is modally remote from now, just as the future is, and the past and the future are in this sense symmetric with respect to now.[21]

Within the mental models framework (Johnson-Laird, 1983), model-based representations are common for modal reasoning with stories (Byrne, 2002). Byrne, Segura, Culhane, Tasso, and Berrocal (2000) provided experimental evidence for model-based representation of temporality in counterfactual scenarios, in which blame and guilt for a bad outcome are attributed to the agent of the most recent event in the scenario, rather than that of the latest-mentioned event. In the same spirit as proponents of situation models, the authors concluded that mental models are “mental representations that are close to the structure of the world rather than to the structure of the language that describes the world” (Byrne et al., 2000, p. 276).

At a more ‘on-line’ level of language processing, there is evidence for shifting between mental models in eye-tracking and ERP studies of counterfactual scenarios (Ferguson & Sanford, 2007; Ferguson, Sanford, & Leuthold, 2008). In a counterfactual

[21] An intuition against Jaszczolt’s (2009) view is that certainty has intuitive endpoints at complete certainty (100%) and complete uncertainty (0%), whereas time has no such endpoints in folk physics. The Big Bang is the closest candidate, but it is only in the past, and there is no uniform intuition about an apocalypse. For discussions of metaphysical time-symmetry, see, e.g., Price (1996).

scenario followed by a real-world violation that is congruent with the counterfactual context (e.g., *It would be great if cats were vegetarian. / If cats were not carnivores* […] *the owner could feed the cat carrots* […]), initial gazes at the critical region (*carrots* […]) were longer and N400s indicating integration difficulty were observed in comparison to regular scenarios (non-counterfactual scenarios without a real-world violation) (Ferguson & Sanford, 2007; Ferguson et al., 2008). In contrast, eye-movement measures in post-critical regions (after *carrots*) were indistinguishable from those in regular scenarios, in fact showing processing difficulty when the subsequent discourse violates the counterfactual context and is consistent with real-world knowledge (*the owner could feed the cat fish*). Ferguson et al. (2008) argued that people rapidly evaluate discourse based on real-world knowledge, and thus real-world violations in the presence of counterfactual context cause an initial disruption in processing that is later neutralized after adopting the counterfactual world.

Following the same reasoning as in the temporal domain, a situation model with magnitude representation predicts modal distance effects, with people exhibiting greater cognitive effort for worlds more remote from the actual world (though not necessarily in a linear relationship). Consistent with this, van Berkum, Hagoort, and Brown (1999) found that the N400 signal is sensitive to discourse incoherence or violations of expectation that do not contain implausibility at the local sentence level (see Hagoort, Hald, Bastiaansen, & Petersson, 2004, for impact of real-world violations in sentences on N400; cf. Kutas & Federmeier, 2000, for an account of N400 effects based on category relations in long-term semantic memory rather than plausibility). Plausibility effects in reading (see Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007) can be captured within a framework of situation models representing modal distances, with the actual world we live in as the default world. Such contextual anticipation based on real-world knowledge of typical situations is consistent with situation models of language meaning across sentences, tracking degrees of plausibility. Because modal distance is determined by subjective epistemic certainty, judgments can often differ between individuals, and conversation is one way of aligning disparate representations (Pickering & Garrod, 2006).

Negation and polarity vary metaphysical distance rather than the narrator's epistemic modality, and thus should not affect modal distance. Evidence indicates that rapid evaluation of a situation model is insensitive to negation, either in a counterfactual scenario (Ferguson et al., 2008) or a non-counterfactual one (see Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007, showing evidence for mental simulation of negated content). To the best of my knowledge, a direct comparison of counterfactual vs. non-counterfactual scenarios involving a negated statement has yet to be conducted. A modal distance account would predict greater difficulty with counterfactuals (involving a modal shift into a distant D-domain) than with non-counterfactuals (suppression of the negated content in the P-domain), but predicting the relative cognitive costs of the suppression in the P-domain vs. the modal shift into a D-domain is not straightforward.

6.3.3.1. Temporal dynamics of *to*-infinitives and actuality.

The pervasiveness of temporal movements in discourse extends even to *to*-infinitives by lexical implicatures of varying degrees of actuality. *To*-infinitives were common in our narratives (235 occurrences in 33 sample narratives), and roughly 30% of the occurrences referred to eventualities that were clearly realized in the actual (story) world.[22] For example, in (4), the topic time does not remain at the time associated with the tensed verb (T1) but moves onto the one associated with the infinitives (T2):

(4) They [Tom's family] were forced$_{T1}$ to leave$_{T2}$ the restaurant and take$_{T2}$ Froggie with them. Tom's family was very annoyed$_{T3}$.
($_T$ for 'time')

In the picture for (4), Tom's family are already outside the restaurant with the frog. The topic time – in the past of the utterance time throughout (4) – moves from T1, when the restaurant manager forces the family to leave, to T2, when the family actually leave with the frog, and then to T3, an interval of the family's annoyance containing T2 by implicature of extended duration for statives (Dowty, 1986). In deciding whether T1 is identical to T2, the entailment or implicature properties of the matrix verb along with the situational context play a critical role. If 'A forcing B to leave' and 'B leaving' are an identical event in Karttunen's (1971) sense because they are spatially and temporally inseparable, we might expect there would be no separate reference time assigned to the *to*-infinitive but simply elaboration of an existing discourse referent (from the matrix verb) with a specified verbal complement (from the infinitive). In whichever way the reader/listener chooses to represent the eventualities, a situation model with rapid tracking of topic time should capture these time movements associated with infinitival

[22] In syntactic literature, control constructions with a complement clause are generally analyzed as having an internal tense interpreted as 'unrealized' or 'quasi-future,' whereas raising constructions have no internal tense operator, thus leaving the temporal interpretation of a *to*-infinitive open to the lexical semantics of the matrix verb (Bresnan, 1972; Stowell, 1982; Martin, 2001; cf. Boeckx & Hornstein, 2006). In a unified account of control and raising such as Boeckx and Hornstein's (2006) movement theory of control, infinitival clauses as well as even some finite clauses (e.g., subjunctives) are 'temporally deficient' in syntax, and there is more room for semantic and pragmatic factors to play a role in temporal interpretation of the complement verb.

verbs as well as with tensed verbs, rather than simply maintain the topic time of the tensed main-clause verb.

Another participant provided the following text for a picture in which a turtle ("Slowpoke") is not in the process of running over to a boy ("Jimmy") but is already in the middle of reporting a frog missing to the boy after running over to him, fulfilling the purpose denoted by the infinitive:

(5) […] Slowpoke ran$_{T1}$ over to Jimmy to show$_{T2}$ him that Fred was missing.

A model of temporal dynamics that is sensitive only to tensed verbs and not to infinitival forms would remain at T1 and miss the temporal contribution from T2 associated with the *to*-infinitive – an undesirable result given the picture. In (5), we see that the combination of situational context and the inherent semantic properties of the matrix verb is crucial to the processing of topic time movement and the contributions of *to*-infinitives (see also Elson & McKeown, 2010, for a perspective from Natural Language Generation on narrative time movement with infinitives such as *to have -ed* vs. *to stop -ing* vs. *to begin -ing*, etc.).

The fact that many of the inferences regarding the actuality of the complement eventualities are not logical entailments would be captured in a rich account of temporal and modal distance representation. So-called 'non-implicative' verbs (e.g., *decided (to* V*)* and *allowed (someone to* V*)*; see Karttunen, 1971) do not commit the speaker to strict truth/falsity regarding the realization of their infinitival complements in the actual world. In such cases, whether topic time advances in the actual-world timeline (P-domain) or branches off into a modally remote timeline (D-domain) would depend on the strength of the implicature from the matrix verb regarding the actualization of its infinitival complement. Although these verbs do not entail the realization of the event in their complement, they usually implicate it, particularly in compatible situational context. If we consider sentences containing such verbs in isolation, we might expect a reader to infer that the event in the complement did not actually happen; if it had, the storyteller would have simply asserted the complement event with a finite verb form (saying *did something* rather than *decided to do something*). On the other hand, given shared access to the target content explicitly showing fulfillment of the eventuality denoted by the complement verb, these non-implicative matrix verbs enrich the discourse with intentions and thematic relations.

Weak-implicating verbs are more likely simply to maintain the topic time or result in temporal progression in a D-domain rather than progression in the P-domain:

(6) […] the waiter tried$_{T1a}$ to grab Freddie [the frog] as he [the frog] looked$_{T1b}$ at the couple. / The waiter caught$_{T2}$ him and picked$_{T3}$ him up by the feet […] ($_{Ta}$ and $_{Tb}$ for parallel actions by multiple agents)

In the picture corresponding to the first text, the waiter is not grabbing the frog yet, so the participant indicates the successful result only in the next text (*the waiter caught him*) to advance reference time.

In sum, the interpretation of *to*-infinitives in discourse also demonstrates the pervasiveness of magnitude representation (in this case, the strength of implicature from a matrix verb to a complement verb) underlying times and worlds in language. These observations are also consistent with coherence-based theories of discourse (e.g., Kehler, 2002; Hobbs, 1990) in that our knowledge of and inferences about eventualities drive temporal interpretation in discourse regardless of the grammatical status (tensed vs. non-tensed) of the verb forms.

6.4. Broader Implications

The situation model framework continues the trend toward broadening the scope of language meaning to include common-sense reasoning based on real-world knowledge (e.g., in the Segmented Discourse Representation Theory; Asher & Lascarides, 2003) and to view language meaning as change in "the information state of anyone who accepts the news" (Veltman, 1996, p. 221; see also Akman, 2009), where sources of information are not limited to explicit verbal information but include perceptual and conceptual access (e.g., bridging inferences, Haviland & Clark, 1974; conceptual anaphors, Gernsbacher, 1991; referential competition in Fukumura et al., 2010; see also Bittner, 2012). Rich representations of discourse including topic time dynamics can improve the performance of automatic discourse segmentation and coreference resolution systems (see Mani, 2010, for an extensive look at narrative computing for time in fiction and poetry); thus, development of time-annotated corpora will be important for training supervised machines (Verhagen et al., 2009). Rich representations can also facilitate automatic generation of natural referring expressions (McCoy & Strube, 1999; van Deemter, Gatt, van Gompel, & Krahmer, 2012) and tenses and aspects (Elson & McKeown, 2010). Sharing insights across related disciplines – psycholinguistics, neurolinguistics, theoretical linguistics, and computational linguistics – will be critical to an adequate account of discourse dynamics and the semantics-pragmatics interface.

Situation models of discourse have practical applications as well. Rapid, automatic use of situation models and shifting between times and worlds at breaks in the story content suggest that more specific and accurate situation models with richer background knowledge of the target content and of typical situations can facilitate language comprehension and organization on standardized tests and in school (see, e.g., Willingham, 2006, and other articles in the Spring 2006 and the Winter 2010-2011 issues of American Educator for support for a knowledge-rich core curriculum). In a recent textbook on cognition, Anderson (2010) reviews the evidence for facilitatory effects of textual elaboration on content recall (e.g., Anderson & Bower, 1973), and specifically suggests elaborative processes such as linking text content to prior knowledge and thinking of specific examples as effective techniques for studying textual material (see also Harley, 2008, Chapter 12).

Appendix

(1) A sample narrative written by a participant in the Frogbook study

"Billy, hurry up!", Billy's mom shouted up the steps. Billy had to get ready. It was his parents' anniversary, and they were going to have a nice dinner. Billy was excited. He loved fancy meals.

Of course, he was always reluctant to leave his pets behind. His dog, Rufus, his turtle, Sheldon, and his frog, Croaker, were his best friends. Of course, they could not come to the restaurant.

Croaker had different ideas though, and he jumped into Billy's sleeve unnoticed.

It was time to go! Billy waved goodbye to his pets, never realizing that Croaker had hitched a ride in his jacket.

The family had arrived at their destination, a wonderful place called "Fancy Restaurant." Everyone was looking forward to their meal inside.

While the family listened to the waiter describe the specials, Croaker took the opportunity to explore. He wasn't interested in eating, so he leaped out of Billy's pocket and went towards the band playing for the evening's entertainment.

But, he landed right into the tuba! It was rather dark.

The tuba player, Alfred, felt some resistance and grew concerned. The other band members also noted that the last note he hit was rather wrong.

Upset, Alfred inspected his tuba, hoping to find the cause of this blockage. He did not want anything to ruin the entertainment for the guests!

"Ack!", Alfred screamed, as Croaker fell out and landed on his face. This knocked Alfred off balance, and he began to trip.

He landed right into the drum! The drum was busted, much to the chagrin of the drummer. Croaker had just ruined the entertainment, but for some, that was far more entertaining!

While the band members argued about exactly what just transpired, Croacker decided to take a trip with the waiter. He jumped into the salad, and enjoyed a ride.

The waiter was, of course, oblivious to this, and went on delivering the salad. It was given to a nice rich woman, who was in for a big surprise.

As she ate, Croacker decided to come out and say hello.

"Croak", he said, in his most charming way.

"AHH!" Screamed the woman. "There's a frog in my salad! Someone, do something!" Croaker could tell he wasn't wanted there, so he hopped away to find some new friends.

And landed with a splash inside a man's glass of wine. The man was engaged in a conversation at the time, and he failed to notice the frog who came to visit.

That is, until he went to have a sip! Croaker came out of the glass to give the man a big kiss on the nose. Croaker was always a loving type.

The man was confused. He had no idea how a frog got in his glass, as he was oblivious to the events that had occurred previously. Croaker liked that he didn't scream, however, and stayed with this new friend.

Until the drummer from before snuck up behind him. Looks like Croaker's fun might be over.

The drummer went to throw Croaker out of the restaurant where he would no longer be a nuisance. Luckily, Billy saw him and was able to shout out, "Wait! That's my frog!"

His family was not pleased by this outburst. They had only just been seated and had not even been served yet. Billy was always so loud.

Billy persisted, and explained that Croaker was his pet and that he did not want to be thrown out. The drummer had a compromise in mind.

He let Billy keep his frog, but kicked the whole family out of the restaurant. They would not be eating there again. Everyone was upset, as they had been very hungry and looking forward to their meals.

The whole drive home, no one spoke to Billy. They all blamed him for what had happened, and Billy began to feel bad. He didn't mean to get everyone kicked out of the restaurant, but still, everyone was angry.

Upon getting home, Billy was immediately sent to his room. He was disappointed, but he went.

Of course, secretly Billy thought what had happened was hilarious, and now back with his best friends, he was able to laugh about it. Croaker was wonderful, and really, despite being hungry, they all had a good time.

(2) Two-Minute Mystery stimuli and sources (all “10 Minutes” versions, with probes)

Story 1: Inspector Turner Nails an Inside Job (*adapted from The Bathroom Readers' Institute, 2008, p. 112*)

It was an inside job. It had to be.

Graying gumshoe Inspector Turner muttered as he lit a cigarette.

The distinguished man beside him, curator of the City Art Museum, eyed him sternly.

Turner crushed the butt on his boot heel.

It was Sunday afternoon, and the museum's most precious work - a Renaissance masterpiece by Bellini - had just been stolen.

Other than the blank spot where Madonna and Child had once hung, there were no clues. The thieves had come and gone during regular museum hours.

Based on the testimony of the security guards, it appeared the picture had disappeared between 11:00 and 11:30 a.m.

Turner was demanding that he interview the four security guards.

10 minutes later, he got the permission to interview them.

(# guards)

The first one claimed he had been patrolling the Egyptian antiquities wing at the time of the heist.

The second one said he had been getting the mail.

The third had been keeping a child from crayoning a Picasso.

And the fourth confessed that he had snuck off to a nearby coffee shop.

10 minutes later, police arrived to arrest the thieves' accomplice that Turner had found.

(# caffeine)

Who was it, and how did he know?

Story 2: It's Only a Matter of Time (*adapted from Silverthorne & Warner, 2010, p. 28*)

It was Joey's 10th birthday party, and he and 7 of his friends were enjoying a scavenger hunt.

The first 9 cues were easy and the boys quickly found the toys hidden throughout the house.

The 10th and final clue read, "At noon and no earlier, you will be shown where to find the next clue."

It was only 11 a.m., and the boys at the birthday party were bummed that they would have to wait a whole hour to continue their treasure hunt.

They were sitting and staring at the clock.

10 minutes later, one of the boys shouted,

(# clock)

"We don't have to wait!

I know where the clue is now!"

10 minutes later, having found the final toy, all the boys were happily eating cake and ice cream and talking excitedly about who had solved the most clues.

(# kids)

How did the boy know where to find the next clue when it wasn't time yet?

Story 3: Ready for Bed (*adapted from Silverthorne & Warner, 2010, p. 81*)

It was 11:30 p.m., and the tourist desperately needed to sleep without any distractions.

He closed the curtains of his hotel window, unplugged his phone, and turned off his alarm.

10 minutes later, he dragged himself out of bed because he realized he had forgotten to put the Do Not Disturb sign on his door.

(# ring)

After putting the sign on his door, he wearily crawled back into bed and began to dream.

10 minutes later, he was suddenly awoken.

(# dream)

Why did he wake up?

Story 4: Fishing (*adapted from Shannon & Sis, 1985, pp. 9-10*)

One fine summer day, two fathers and two sons awoke at 6 a.m. to go fishing at their favorite lake.

10 minutes later, they had their fishing poles and worms out.

(# river)

They fished and talked all morning long and by noon everyone had caught one fish, so they started celebrating and drinking beer.

10 minutes later, the two fathers and two sons walked back home. Everyone was happy because each had a fish even though only three fish had been caught.

(# beer)

There were two fathers and two sons, but only three fish were caught and no fish were lost.

How can this have happened?

Story 5: The Clever Bride (*adapted from Shannon & Sis, 1985, pp. 23-25*)

There was a bride who lived with her mother-in-law and was very fond of chickpeas.

The bride liked them so much that she would steal 15 chickpeas from the kitchen every day to roast and eat in secret while her mother-in-law was tending the plants in the garden.

10 days later, half the sack of chickpeas was gone and the mother-in-law was angry.

(# beans)

She suspected the bride and mumbled to herself,

"I'm certain she's the thief.

She's the only new person in the house."

The mother-in-law was a smart woman, but the young bride was even cleverer.

She knew her mother-in-law suspected her.

One sunny morning, the bride was cleaning house with her mother-in-law.

10 minutes later, the bride found a chickpea on the floor.

(# house)

She picked it up, showed it to her mother-in-law, and said three words that convinced the older woman that she hadn't taken the chickpeas.

What did the bride say?

Story 6: Heaven and Hell (*adapted from Shannon & Sis, 1985, pp. 51-53*)

People are always wishing.

But once in China a man in his 60s got his wish, which was to see the difference between heaven and hell before he died.

When he visited hell, he saw 10 tables crowded with delicious food, but everyone was hungry and angry.

(# deathbed)

The people had food, but they were forced to sit several feet from the table and use chopsticks 3 feet long.

They kept trying in vain to get food into their mouths.

10 minutes later, the man saw heaven, and he was very surprised for it looked the same.

(# food)

Again, big tables of delicious food covered tables, but people were forced to sit several feet from the table and use 3-foot-long chopsticks that made it impossible to get any food into their mouths.

It was exactly like hell, but in heaven the people were well fed and happy.

Why?

Story 7: Inspector Saunders and the Stranger (*adapted from The Bathroom Readers' Institute, 2008, p. 12*)

At 7 o'clock in the evening, Inspector Saunders was sitting in his hotel room in Cleveland, cleaning and oiling his gun.

10 minutes later, he heard someone knock on the door and then insert a key card into the lock.

(# gun)

Inspector Saunders went over and opened the door.

"Yo, sorry, my man!" said the stranger as he took in all of Saunders' 6 feet, 6 inches.

"I thought this was my room.

I must have gotten off on the wrong floor.

They all look identical, don't they?"

The stranger laughed weakly and gave a little wave of the hand as he made his way to the elevator.

The inspector closed the door and went right to the phone to alert the front desk that a thief was stalking the halls.

10 minutes later, police arrested the stranger for trespassing.

(# lounge)

How did Saunders know the stranger was up to no good?

Story 8: Inspector Martin Solves a Kidnapping Case (*adapted from The Bathroom Readers' Institute, 2008, p. 19*)

Tom Travail phoned Inspector Martin at 1 a.m. because he didn't want to call the police.

10 minutes later, Tom told Inspector Martin exactly what had happened.

(# officer)

His son Tony had been kidnapped, and the midnight drop of half a million dollars hadn't gone as planned.

Tony was still missing and his twin, Terry - who had delivered the gym bag full of cash - was nursing a bump on the head.

10 minutes later, Terry told the inspector,

(# bump)

"I went to the deserted parking garage just like they told me.

I was standing there, and all of a sudden somebody hit me on the back of the head.

I fell and dropped the gym bag.

My attacker swooped down from behind me, picked it up, and ran off.

I never saw his face, only his back.

He was tall and red-headed, wearing chinos and a zippered sweat shirt - it might have had a college logo, I couldn't quite see in the dark."

"Now they want another half million," Tom wailed.

"What should we do?"

"I know what happened," said Martin.

How and what did Inspector Martin know?

Story 9: Inspector Williams Solves the "Lipstick on His Collar" Case (*adapted from The Bathroom Readers' Institute, 2008, p. 85*)

Joe Jones had been murdered on his front porch, his key still in the lock and his wallet missing.

Dressed in a business suit, he had apparently been leaving for work when he was assaulted around 7:30 a.m. - or so the police assumed.

10 minutes later, Inspector Williams arrived at the scene of crime.

(# commute)

Williams noticed a red lipstick smear on Joe's shirt collar.

Surely Joe wouldn't report to the office like that.

Williams was thinking that Joe was probably coming in from a night at his girlfriend's when someone shot him.

10 minutes later, the cops invited Inspector Williams to interview the witnesses.

(# night)

A taxi driver remembered dropping someone off a couple of blocks away, and he heard what sounded like a car backfiring around 7:30 a.m.

A woman walking her dog in a nearby park also heard a sound, but didn't see anything.

Nor did a mailman who had been in the vicinity.

"Poor guy," the mailman said.

"Someone must have seen him coming home and attacked him on his porch.

But I was in my truck sorting mail - I didn't hear the shot."

Inspector Williams then advised the cops to check with Jones's girlfriend:

"If he spent the night at her place, then we have our killer."

Who did Williams suspect and why?

Story 10: In Hot Water (*adapted from Silverthorne & Warner, 2010, p. 18*)

A bank robber jumped into his car and was speeding away.

10 minutes later, he rushed into his house and started the bath water.

(# speeding)

He undressed, splashed some water on himself, and put on a bathrobe.

Just as the tub was overflowing, he heard police pounding on the front door.

He knew the police suspected him of robbing the bank, and he needed an alibi.

10 minutes later, he was taken to the police station under arrest.

(# crook)

How did the police figure out that the bank robber hadn't taken a bath?

Story 11: Murder at the Inn (*adapted from The Bathroom Readers' Institute, 2008, p. 184*)

Trouble seemed to follow Inspector Matthews - even on vacation.

He was on a remote island of the Florida Keys, hoping to do some deep-sea fishing, when his vacation was interrupted by a murder at the inn where he was staying.

At 12:15 a.m. Matthews was resting in his room after downing a few Guinnesses at a nearby bar.

10 minutes later, the proprietors phoned him to ask for his help.

(# bar)

"I heard the shot at 12:15 a.m.," Mrs. Barlow told him.

"I was sitting at my vanity table brushing my hair and putting on moisturizer."

Her husband nodded and said,

"And I was in bed reading."

"Are you sure of the time?" the inspector asked.

He had left the bar at midnight and was back in his room shortly thereafter.

A shot at 12:15 would have grabbed his attention.

But he hadn't heard one.

"Positive," Mrs. Barlow said.

"I was so shocked that I looked up from what I was doing, glanced up and distinctly saw the face of our grandfather clock reflected in the vanity mirror.

It clearly read 12:15."

10 minutes later, distant police sirens sounded in the night.

(# grandmother)

"The cops are here," Inspector Matthews said.

"Now I know what time to tell them the shot was fired."

What time did Mrs. Barlow actually hear the shot?

Story 12: The Night Shift (*adapted from The Bathroom Readers' Institute, 2008, p. 134*)

The night watchman at a factory settled into his chair after eating a heavy meal.

He was watching his favorite late night show on television.

10 minutes later, he was fast asleep.

(# television)

At 8 a.m. the next morning, his boss arrived, and the watchman told his boss about a dream he'd had:

Someone was planning to sabotage the factory.

The watchman was telling his boss it was an omen.

10 minutes later, the boss fired him.

(# warning)

Why?

Story 13: A Drink for a Crow (*adapted from Shannon & Sis, 1985, pp. 17-19*)

Once there was a crow who had grown so thirsty he could barely caw.

He flew down to a tall water pitcher where he had gotten a drink of water 3 days ago, but there was only a little bit of water remaining at the bottom.

He tried and tried to reach it with his beak, but the pitcher was too deep and his beak was too short.

10 minutes later, he figured out what to do.

(# wing)

He kept flying back and forth from the pebble beach to the pitcher.

10 minutes later, he was able to drink easily from the pitcher while sitting on its edge.

(# beach)

What did the crow do?

Story 14: Which Flower? (*adapted from Shannon & Sis, 1985, pp. 20-22*)

Once long ago, the Queen of Sheba decided to test King Solomon's wisdom by a series of 7 tests and riddles.

He passed each one with ease.

10 minutes later, the queen led him to a room filled with flowers of every shape and color.

(# puzzle)

The finest craftsmen and artists had constructed the flowers so that they looked exactly like the real flowers from her garden.

"The test," she told King Solomon, "is to find the one real flower among all the artificial ones."

King Solomon carefully looked from flower to flower and back again, searching for even the smallest of differences.

He looked for any sign of wilted leaves or petals, but found lifelike leaves and petals in all conditions on every flower.

And fragrance was of no help, for the room was filled with fragrances, so King Solomon was looking for other clues.

10 minutes later, he decided to make a suggestion:

(# fragrance)

"Please," said King Solomon. "This room is so warm.

Could we open the window and let in the breeze?

The fresh air will clear my head for thinking."

The Queen of Sheba kindly agreed, and within minutes after the window had been opened King Solomon knew which of the many was the one real flower.

How did he suddenly know?

Story 15: The Sticks of Truth (*adapted from Shannon & Sis, 1985, pp. 26-29*)

Long ago in India judges traveled from village to village.

One day a judge stopped at an inn to rest, but the innkeeper was very upset.

Someone had just that day stolen his daughter's gold ring.

The judge told him not to worry and had all 15 guests gather so that he could question them.

He could not figure out from their answers who the thief was, so he was getting frustrated.

10 minutes later, the judge decided to use some old magic.

(# frustrated)

He told them all he was going to have to use the sticks of truth.

"These are magic sticks," he explained, "that will catch the thief."

He gave each guest a stick to put under their bed during the night.

"The stick belonging to the thief will grow during the night.

At breakfast we will all compare sticks and the longest stick will be the thief's."

The next morning the judge had all the guests gather in the dining room.

10 minutes later, he asked them to come by his table and hold their sticks up next to his to see if they had grown.

(# napkin)

But one after another all were the same.

None of them had grown any longer.

Then suddenly the judge called,

"This is the thief! Her stick is shorter than all the rest."

Once caught, the woman confessed and the ring was returned.

But all the guests were confused about the sticks of truth.

The judge had said the longest stick would be the thief's, but instead it had been the shortest stick.

Why?

References

Akman, V. (2009). Situated semantics. In P. Robbins, & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 401-418). New York: Cambridge University Press.

Aksu-Koç, A. A., & von Stutterheim, C. (1994). Temporal relations in narrative: Simultaneity. In R. A. Berman, & D. I. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

American Federation of Teachers. (Spring 2006). *American Educator, 30*(1).

American Federation of Teachers. (Winter 2010-2011). *American Educator, 34*(4).

Anderson, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior, 13*(2), 149-162.

Anderson, J. R. (2010). *Cognitive psychology and its implications* (7th ed.). New York: Worth Publishers.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, D.C.: V. H. Winston.

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative texts. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35*(A), 427-440.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56*, 521-536.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. New York: Cambridge University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Bagga, A., & Baldwin, B. (1998). Algorithms for scoring coreference chains. In *Proceedings of LREC Workshop on Linguistic Coreference* (pp. 563-566).

Barr, D. J., & Gann, T. M. (2011). *Audience design as expert performance*. Paper presented at Pre-CogSci 2011: Bridging the gap between computational, empirical and theoretical approaches to reference, Boston, MA.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: The MIT Press.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bestgen, Y. (1998). Segmentation markers as trace and signal of discourse structure. *Journal of Pragmatics, 29*, 753-763.

Bestgen, Y., & Vonk, W. (2000). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language, 42*, 74-87.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.

Bird, S., Loper, E., & Klein, E. (2011). The Natural Language Toolkit (Version 2.0b9). Available from http://www.nltk.org/download

Bittner, M. (1999). Concealed causatives. *Natural Language Semantics, 7*, 1-78.

Bittner, M. (2011). Time and modality without tenses or modals. In R. Musan, & M. Rathert (Eds.), *Tense across languages* (pp. 147-188). Niemeyer, Tübingen: de Gruyter.

Bittner, M. (in preparation). *Temporality: Universals and variation*. Wiley-Blackwell.

Boeckx, C., & Hornstein, N. (2006). The virtues of control as movement. *Syntax, 9*(2), 118-130.

Bohnemeyer, J. (2009). Temporal anaphora in a tenseless language. In W. Klein, & P. Li (Eds.), *The expression of time* (pp. 83-128). Berlin: Mouton de Gruyter.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*, 1-22.

Botne, R. (2012). Remoteness distinctions. In R. I. Binnick (Ed.), *The Oxford Handbook of Tense and Aspect* (pp. 536-563). New York: Oxford University Press.

Botne, R., & Kershner, T. L. (2008). Tense and cognitive space: On the organization of tense/aspect systems in Bantu languages and beyond. *Cognitive Linguistics, 19*(2), 145-218.

Brannon, E. M., & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing* (pp. 143-182). Boca Raton, FL: CRC Press.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Brennan, S. (1995). Centering attention in discourse. *Language and Cognitive Processes, 10*, 137-167.

Bresnan, J. (1972). *Theory of complementation in English syntax*. Doctoral dissertation, MIT.

Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 1-59). Cambridge, MA: The MIT Press.

Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences, 6*(10), 426-431.

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory & Cognition, 28*(2), 264-281.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition, 25*(4), 438-446.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning, 58*: Supplement s1, 63-79.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C. Li (Ed.), *Subject and topic* (pp. 25-55). New York: Academic Press.

Chafe, W. L. (1980). *The pear stories: Cognitive, cultural and linguistic aspects of narrative production. Vol. 3. Advances in discourse processes*. Norwood, NJ: Ablex. (Now available from Westport, CT: Greenwood Press).

Chen, J.-Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104*, 427-436.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 127-202). Norwood, NJ: Ablex.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam: North Holland Publishing.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition, 7*(1), 35-41.

Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.

Dahl, Ö. (1984). Temporal distance: Remoteness distinctions in tense-aspect systems. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Explanations for Language Universals* (pp. 105-122). New York: Mouton.

Dahl, Ö. (1985). *Tense and aspect systems*. Oxford: Blackwell.

van Deemter, K., Gatt, A., van Gompel, R. P. G., & Krahmer, E. (2012). Towards a computational psycholinguistics of reference production. *Topics in Cognitive Science, 4*(2), 166-183.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*, 145–147.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener." In D. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics* (pp. 105-129). New York: Academic Press.

Denis, P., & Baldridge, J. (2008). Specialized models and ranking for coreference resolution. In *Proceedings of EMNLP 2005* (pp. 660-669).

Devlin, K. (1990). Infons and types in an information-based logic. In R. Cooper, K. Mukai, & J. Perry (Eds.), *Situation theory and its applications: Volume 1* (pp. 79-96). Center for the Study of Language and Information. Stanford, CA.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2008). Time travel through language: Temporal shifts rapidly decrease information accessibility during reading. *Psychonomic Bulletin & Review, 15*(4), 750-756.

Dowty, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht, Holland: D. Reidel Publishing Company.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy, 9*(1), 37-62.

Efstathiadi, L. (2009). *The use of epistemic modality markers as a means of hedging and boosting by L1 and L2 speakers of Modern Greek: A corpus-based study in informal letter-writing*. Doctoral dissertation, Aristotle University of Thessaloniki.

Elson, D. K., & McKeown, K. R. (2010). Tense and aspect assignment in narrative discourse. In *Proceedings of INLG 2010*. Dublin, Ireland.

Ferguson, H. J., & Sanford, A. J. (2007). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language, 58*(3), 609-626.

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research, 1236*, 113-125.

Franklin, M. S., Smith, E. E., & Jonides, J. (2007). Distance effects in memory for sequences: Evidence for estimation and scanning processes. *Memory, 15*(1), 104-116.

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science, 34*, 1430-1451.

Fukumura, K., van Gompel, R. P. G., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *The Quarterly Journal of Experimental Psychology, 63*(9), 1700-1715.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59-65.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Philadelphia, PA: Psychology Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A. (1991). Comprehending conceptual anaphors. *Language and Cognitive Processes, 6*(2), 81-105.

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics, 30*, 5-55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of statements about story worlds that deviate from normal conceptions of time: What is true about *Einstein's Dreams*? *Cognitive Psychology, 35*(3), 246-301.

Grimes. J. E. (1975). *The thread of discourse*. The Hague: Mouton.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274-307.

Haberlandt, K. F., Berian, C., & Sandson, J. (1980). The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior, 17*, 419-425.

Haghighi, A., & Klein, D. (2007). Unsupervised coreference resolution in a nonparametric Bayesian model. In *Proceedings of ACL 2007*, pages 848–855.

Haghighi, A., & Klein, D. (2009). Simple coreference resolution with rich syntactic and semantic features. In *Proceedings of EMNLP 2009* (pp. 1152–1161). Singapore.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-441.

Hamm, F., Kamp, H., & van Lambalgen, M. (2006). There is no opposition between formal and cognitive semantics. *Theoretical Linguistics, 32*(1), 1-40.

Harley, T. A. (2008). *The psychology of language: From data to theory* (3rd ed.). New York: Psychology Press.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring New information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 512-521.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Malden, MA: Blackwell Publishers.

Hinrichs, E. W. (1988). Tense, quantifiers, and contexts. *Computational Linguistics, 14*(2), 3-14.

Hobbs, J. (1990). *Literature and cognition*. Stanford, CA: Center for the Study of Language and Information.

Holmes, E. T., & Tudor, T. (1977). *Amy's goose*. New York: HarperTrophy.

Jackendoff, J. (1978). Grammar as evidence for conceptual structure. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 201-228). Cambridge, MA: The MIT Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Jansen, C. J. M., & Pollmann, M. M. W. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics, 8*(3), 187-201.

Jaszczolt, K. M. (2009). *Representing time: An essay on temporality and modality*. New York: Oxford University Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to Natural Language Processing, computational linguistics, and speech recognition.* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kamp, H. (1979). Events, instants and temporal reference. In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 376-417). Berlin: Springer-Verlag.

Kamp, H. (1981). *Discourse representation and temporal reference*. (French translation published as Evénements, représentations discursives et référence temporelle. *Langages, 64*(15), 39–64).

Kamp, H. (1990). Prolegomena to a structural account of belief and other attitudes. In C. Anderson, & J. Owens (Eds.), *Propositional attitudes: The role of content in logic* (pp. 27-90). Stanford, CA: Center for the Study of Language and Information.

Kamp, H. & Reyle, U. (2008). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Publishers.

Kantola, L., & van Gompel, R. P. G. (2011). *Does the address matter when choosing referring expressions?* Paper presented at Pre-CogSci 2011, Boston, MA.

Karttunen, L. (1971). Implicative verbs. *Language, 47*(2), 340-358.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology, 60*, 976-990.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: Center for the Study of Language and Information.

Kelter, S., Kaup, B., & Claus, B. (2004). Representing a described sequence of events: A dynamic view of narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 451-464.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Klein, W. (1994). *Time in language*. New York: Routledge.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Lapata, M., & Lascarides, A. (2006). Learning sentence-internal temporal relations. *Journal of Artificial Intelligence Research, 27*, 85-117.

Lee, C. (2012). *Situation model and salience*. Paper presented at The LSA 2012 Special Session on Information Structure and Discourse: In Memory of Ellen F. Prince. Portland, Oregon.

Lee, C., Kharkwal, G., & Stromswold, K. (2012). Temporal transitions in narrative production with wordless picture books. *eLanguage.net: LSA Meeting Extended Abstracts 2012*.

Lee, C., Muresan, S., & Stromswold, K. (2012). *Computational analysis of referring expressions in narratives of picture books*. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Montréal, Canada.

Lee, C., & Stromswold, K. (submitted). Situation model and accessibility: Referring expressions in narrative production.

Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., & Jurafsky, D. (2011). Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 Shared Task. In *Proceedings of the CoNLL-2011 Shared Task* (pp. 28-34).

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*(1), 339-359.

Li, W. (2004). Topic chains in Chinese discourse. *Discourse Processes, 37*(1), 25-45.

Litteral, R. (1972). Rhetorical predicates and time topology in Anggor. *Foundations of Language, 8*, 391-410.

Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., & Roukos, S. (2004). A mention-synchronous coreference resolution algorithm based on the Bell tree. In *Proceedings of ACL 2004* (pp. 135-142).

Madden, C. J. & Zwaan, R. A. (2003). How does verb aspect constrain event representation? *Memory & Cognition, 31*, 663-672.

Magliano, J. P. & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse processes, 29*, 83-112.

Magliano, J. P., Zwaan, R. A., & Graesser, A. (1999). The role of situational continuity in narrative understanding. In H. van Oostendorp, & S. R. Goldman, (Eds.), *The construction of mental representations during reading* (pp. 319-340). Mahwah, NJ: Lawrence Erlbaum Associates.

Malaka, R., & Zipf, A. (2000). Deep Map: Challenging IT research in the framework of a tourist information system. In D. R. Fesenmaier, S. Klein, & D. Buhalis (Eds.), *Information and Communication Technologies in Tourism 2000: Proceedings of the International Conference in Barcelona, Spain* (pp. 15-27). Wien: Springer.

Mani, I. (2010). *The imagined moment: Time, narrative, and computation*. Lincoln, NE: University of Nebraska Press.

Marchman, V. (1989). *Episodic structure and the linguistic encoding of events in narrative: A study of language acquisition and performance*. Doctoral dissertation, University of California, Berkeley.

Martin, R. (2001). Null Case and the distribution of PRO. *Linguistic Inquiry, 32*(1), 141-166.

Mayer, M. (1969). *Frog, where are you?* New York: Penguin Books.

Mayer, M. (1974). *Frog goes to dinner*. New York: Penguin Books.

Mayer, M, & Mayer, M. (1975). *One frog too many*. New York: Penguin Books.

McCoy, K. F., & Strube, M. (1999). Taking time to structure discourse: Pronoun generation beyond accessibility. In M. Hahn & S. Stoness (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 378-383). Mahwah, NJ: Lawrence Erlbaum Associates.

McKoon, G. & Ratcliff, R. (1980). The comprehension processes and memory structures involved anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19*, 668-682.

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics, 14*(2), 15-28.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13*(2), 180-184.

Mozuraitis, M., Chambers, C., & Daneman, M. (2011). *Online integration of verb aspect and world knowledge during reading*. Paper presented at CUNY 2011. Stanford, CA.

Müller, C., & Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. In Sabine Braun, Kurt Kohn, and Joybrato Mukherjee (Eds.), *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods* (pp. 197-214). Frankfurt: Peter Lang.

Nairn, R., Condoravdi, C., & Karttunen, L. (2006). Computing relative polarity for textual inference. In *Proceedings of Inference in Computational Semantics (ICoS-5)*. Buxton, UK.

Ng, V. (2008). Unsupervised models for coreference resolution. In *Proceedings of EMNLP 2008* (pp. 640-649).

Ng, V. (2010). Supervised noun phrase coreference research: The first fifteen years. In *Proceedings of ACL 2010* (pp. 1396-1411).

Ng, V., & Cardie, C. (2002). Improving machine learning approaches to coreference resolution. In *Proceedings of ACL 2002* (pp. 104-111).

O'Connor, B., & Heilman, M. (2011). ARKref is a Noun Phrase Coreference System. Website at http://www.ark.cs.cmu.edu/ARKref/

Partee, B. H. (1984). Nominal and temporal anaphors. *Linguistics and Philosophy, 7*(3), 243-286.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89-110.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*, 203-228.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105-110.

Poon, H., & Domingos, P. (2008). Joint unsupervised coreference resolution with Markov logic. In *Proceedings of EMNLP 2008* (pp. 650-659).

Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Weischedel, R., & Xue, N. (2011). CoNLL-2011 Shared Task: Modeling unrestricted coreference in OntoNotes. In *Proceedings of CoNLL 2011*.

Price, H. (1996). *Time's arrow and Archimedes' point: New directions for the physics of time*. New York: Oxford University Press.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-256). New York: Academic Press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413-425.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416-426.

Rahman, A., & Ng, V. (2009). Supervised models for coreference resolution. In *Proceedings of EMNLP 2009* (pp. 968-977).

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition, 28*(8), 1310-1320.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437-442.

Sanders, T. J. M., Spooren, W. P. M., Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes, 15*, 1-35.

Sanford, A. J., Moar, K., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and Speech, 31*(1), 43-56.

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review, 14*(3), 512-516.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60*, 205-234.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Oxford: Lawrence Erlbaum Associates.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Shannon, G., & Sis, P. (1985). *Stories to solve: Folktales from around the world*. New York: HarperTrophy.

Silverthorne, S., & Warner, J. (2010). *Mind-boggling one-minute mysteries and brain teasers*. Eugene, OR: Harvest House.

Smith, C. (1991). *The parameter of aspect*. Dordrecht: Kluwer Academic Press.

Soon, W. M., Ng, H. T., & Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics, 27*(4), 521-544.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53*, 125-140.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1162-1169.

Stone, M. (1997). The anaphoric parallel between modality and tense. Technical Report IRCS 97-6. Retrieved January 20, 2012, from http://repository.upenn.edu/ircs_reports/79.

Stowell, T. (1982). The tense of infinitives. *Linguistic Inquiry, 13*(3), 561-570.

Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.

Tenbrink, T. (2007). *Space, time, and the use of language: An investigation of relationships*. New York: Mouton de Gruyter.

The Bathroom Readers' Institute. (2008). *Uncle John's bathroom puzzler*. Canada, Portable Press.

Tonhauser, J. (2011). Temporal reference in Paraguayan Guaraní, a tenseless language. *Linguistics and Philosophy, 34*, 257-303.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*, 515-557.

Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic, 25*, 221-261.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Moszkowicz, J., & Pustejovsky, J. (2009). The TempEval challenge: Identifying temporal relations in text. *Language Resources and Evaluation, 43*, 161-179.

Versley, Y., Ponzetto, S. P., Poesio, M., Eidelman, V., Jern, A., Smith, J., Yang, X., & Moschitti, A. (2008). BART: A modular toolkit for coreference resolution. In *Companion Volume of the Proceedings of ACL 2008* (pp. 9-12).

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., & Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Message Understanding Conference* (pp. 45-52).

Vonk, W., Hustinx, L. G. M. M., & Simons, W. H. G. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes, 7*(3/4), 301-333.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483-488.

Wanner, H. E. (1968). *On remembering, forgetting and understanding sentences: A study of the deep structure hypothesis*. Unpublished doctoral dissertation, Harvard University.

Webber, B. (1988). Tense as discourse anaphor. *Computational Linguistics, 14*(2), 61-73.

Weger, U., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review, 15*(2), 426-430.

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology, 2*, 1-11.

Willingham, D. T. (2006). How knowledge helps: It speeds and strengthens reading comprehension, learning – and thinking. *American Educator, 30*(1), 30-37.

Yang, X., Su, J., Zhou, G., & Tan, C. L. (2004). An NP-cluster based approach to coreference resolution. In *Proceedings of COLING* (pp. 226-232).

Yang, X., Zhou, G., Su, J., & Tan, C. L. (2003). Coreference resolution using competitive learning approach. In *Proceedings of ACL 2003* (pp. 176-183).

Yap, F. H., Chu, P. C. K., Yiu, E. S. M., Wong, S. F., Kwan, S. W. M., Matthews, S., Tan, L. H., Li, P., & Shirai, Y. (2009). Aspectual asymmetries in the mental representation of events: Role of lexical and grammatical aspect. *Memory & Cognition, 37*(5), 587-595.

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2011). *Influence of perspective and goals on reference production in conversation.* Paper presented at Pre-CogSci 2011, Boston, MA.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty first annual conference of the cognitive science society* (pp. 778–783). Mahwah, NJ: Erlbaum.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1196-1207.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science, 8*(1), 15-18.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35-62). New York: Academic Press.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297.

Zwaan, R. A., Taylor, L. J., & de Boer, M. (2010). Motor resonance as a function of narrative time: Further tests of the linguistic focus hypothesis. *Brain & Language, 112*, 143-149.